The Annals of Applied Probability
2015, Vol. 25, No. 5, 2563–2625
DOI: 10.1214/14-AAP1056

TIME DISCRETIZATION OF FBSDE WITH POLYNOMIAL GROWTH DRIVERS AND REACTION–DIFFUSION PDES[1]

BY ARNAUD LIONNET[2], GONÇALO DOS REIS[3] AND LUKASZ SZPRUCH

University of Oxford, University of Edinburgh and CMA/FCT/UNL, and University of Edinburgh

In this paper, we undertake the error analysis of the time discretization of systems of Forward–Backward Stochastic Differential Equations (FBSDEs) with drivers having polynomial growth and that are also monotone in the state variable.

We show with a counter-example that the natural explicit Euler scheme may diverge, unlike in the canonical Lipschitz driver case. This is due to the lack of a certain stability property of the Euler scheme which is essential to obtain convergence. However, a thorough analysis of the family of θ-schemes reveals that this required stability property can be recovered if the scheme is sufficiently implicit. As a by-product of our analysis, we shed some light on higher order approximation schemes for FBSDEs under non-Lipschitz condition. We then return to fully explicit schemes and show that an appropriately tamed version of the explicit Euler scheme enjoys the required stability property and as a consequence converges.

In order to establish convergence of the several discretizations, we extend the canonical path- and first-order variational regularity results to FBSDEs with polynomial growth drivers which are also monotone. These results are of independent interest for the theory of FBSDEs.

1. Introduction. There is currently a long literature on the numerical approximation of FBSDE with Lipschitz conditions [Bouchard and Touzi (2004), Crisan and Manolarakis (2012), Gobet and Turkedjiev (2011), Chassagneux (2012, 2013) and references within]. In this article, we address the case of FBSDEs with drivers having polynomial growth in the state variable, which has not been studied before, and provide customized analysis of various implicit and explicit schemes.

Received September 2013; revised June 2014.

[1]Supported by the Engineering and Physical Sciences Research Council (UK) which has funded this work as part of the Numerical Algorithms and Intelligent Software Centre (NAIS) under Grant EP/G036136/1.

[2]Supported in part by the Engineering and Physical Sciences Research Council (UK) via the Grant EP/P505216/1, as well as by the Oxford-Man Institute.

[3]Supported by the *Fundação para a Ciência e a Tecnologia* (Portuguese Foundation for Science and Technology) through the project PEst-OE/MAT/UI0297/2014 (Centro de Matemática e Aplicações CMA/FCT/UNL).

MSC2010 subject classifications. Primary 65C30, 60H35; secondary 60H07, 60H30.

Key words and phrases. FBSDE, monotone driver, polynomial growth, time discretization, path regularity, calculus of variations, numerical schemes.

The importance of FBSDEs with nonlinear drivers is due to the fruitful connection between FBSDEs and partial differential equations (PDEs). Many biological and physical phenomena are modeled using PDEs of parabolic type, say for $(t,x) \in [0,T] \times \mathbb{R}^d$

$$-\partial_t v(t,x) - \mathcal{L}v(t,x) - f\big(t,x,v(t,x),(\nabla v \sigma)(t,x)\big) = 0, \qquad v(0,x) = g(x),$$

with $\mathcal{L}$ a second-order elliptic differential operator and certain measurable functions f and g. A very large class of such equations can be linked to the solution process $\Theta^{t,x} = (X^{t,x}, Y^{t,x}, Z^{t,x})$ of certain forward–backward stochastic differential equations (FBSDEs) with the following type of dynamics for $(t,x) \in [0,T] \times \mathbb{R}^d$, $s \in [t,T]$ and W a Brownian-motion

$$X_s^{t,x} = x + \int_t^s b\big(r, X_r^{t,x}\big)\,\mathrm{d}r + \int_t^s \sigma\big(r, X_r^{t,x}\big)\,\mathrm{d}W_r, \tag{1.1}$$

$$Y_s^{t,x} = g\big(X_T^{t,x}\big) + \int_s^T f\big(r, \Theta_r^{t,x}\big)\,\mathrm{d}r - \int_s^T Z_r^{t,x}\,\mathrm{d}W_r, \tag{1.2}$$

via the so-called nonlinear Feynman–Kac formula: $v(T-t,x) = Y_t^{t,x}$ [see, e.g., El Karoui, Peng and Quenez (1997)].

In many applications of interest, like reaction–diffusion type equations, the function f is a polynomial (in v), for example, the Allen–Cahn equation, the FitzHugh–Nagumo equations (with or without recovery) or the standard nonlinear heat and Schrödinger equation [see Henry (1981), Rothe (1984), Estep, Larson and Williams (2000), Kovács (2011) and references].

Motivated by these applications, we look further at the connection between parabolic PDEs and FBSDEs with monotone drivers f of polynomial growth [see Pardoux (1999), Briand and Carmona (2000) and Briand et al. (2003)]. By monotonicity, we mean that $\langle v' - v, f(v') - f(v)\rangle \le \mu |v' - v|^2$, for some $\mu \ge 0$, and any v, v' (one can also find the terminology that f is one-sided Lipschitz). We extend the above mentioned works by providing further regularity estimates for the FBSDE in question (modulus of continuity, path and variational regularity). Then we proceed to a thorough analysis of various numerical methods that open the door to Monte Carlo methods for solving numerically the corresponding PDEs.

The work and results we present should be understood as a first step in the numerical analysis of FBSDE with monotone drivers of polynomial growth, wider than the Lipschitz driver BSDE setting, with the intent of deepening the applicability of FBSDEs to reaction–diffusion equations. Moreover, we work without assuming knowledge on the density function or the moment generating function of the forward process X. In some applications where X is simply the Brownian motion, it is possible to derive a numerical solver that takes advantage on this knowledge; see, for example, Zhang, Gunzburger and Zhao (2013). The work we develop aims at black-box type algorithms which do not take advantage of any of the specific forms the FBSDEs coefficients may take.

A motivating example. To better understand why the explicit Euler scheme seems not to be suitable for approximating the solution to BSDEs with non-Lipschitz drivers, let us consider the following simple example (for further details and notational setup, see Section 2 and Appendix A.1):

$$Y_t = \xi - \int_t^1 Y_s^3 \, \mathrm{d}s - \int_t^1 Z_s \, \mathrm{d}W_s, \qquad t \in [0, 1] \tag{1.3}$$

with the terminal condition $\xi \in \mathcal{F}_1$. For any $\xi \in L^p$ for $p \geq 2$, there exists[4] a unique (square-integrable) solution (Y, Z) to the above BSDE.

Fix the number of time-discretization points to be $N + 1 > 0$. The explicit Euler scheme for the above equation with uniform time step $h = 1/N$ is, with the notation $Y_i := Y_{i/N}$, given by

$$Y_i = \mathbb{E}\big[Y_{i+1} - Y_{i+1}^3 h | \mathcal{F}_i\big] = \mathbb{E}\big[Y_{i+1}\big(1 - hY_{i+1}^2\big) | \mathcal{F}_i\big], \qquad i = 0, \ldots, N-1, \tag{1.4}$$

where $Y_N = \xi$.

It is a simple calculation (see Appendix A.1 for the details) to show that if

$$\xi \geq 2\sqrt{N} \qquad \text{then } |Y_i| \geq 2^{2^{N-i}} \sqrt{N} \qquad \text{for } i = 0, \ldots, N. \tag{1.5}$$

With this simple computation in mind, it is possible to show that there exists a random variable ξ whose moments of any order are finite and for which the explicit Euler scheme diverges. The result below is a corollary of Lemma A.2 that can be found in Appendix A.1.

LEMMA 1.1. *Let π^N be the uniform grid over the interval* $[0, 1]$ *with $N + 1$ points, N an even number ($t = 1/2$ is common to all grids π^N). For any $\xi \in L^p(\mathcal{F}_1)$, for $p \geq 2$, let (Y, Z) denote the solution to* (1.3).

Then there exists a random variable $\xi \in L^p \setminus L^\infty$ for any $p \geq 2$ such that

$$\lim_{N \to \infty} \mathbb{E}\big[\big|Y_{1/2}^{(N)}\big|\big] = +\infty,$$

where $Y_{1/2}^{(N)}$ is the Euler approximation of Y on the time point $t = 1/2$ via (1.4) *over the grids π^N.*

The special random variable ξ we work with is normally distributed and it is known that $\mathbb{P}[|\xi| > 2\sqrt{N}]$ is exponentially small (see Lemma A.1). What our counter-example shows is that although ξ may take very large values on an event with exponentially small probability, the impact of these very large values when propagated through the Euler explicit scheme is doubly-exponential [see (1.5)].

[4]Existence and uniqueness follows from Section 2 in Pardoux (1999) or Theorem 2.2 below.

This double-exponential impact is precisely a consequence of the superlinearity of the driver. In general, the terminal condition ξ is an unbounded random variable (RV) so there is a positive probability of the scenario where $\xi \geq 2\sqrt{N}$ no matter how small a time-step we choose. This indicates that, in general, the explicit Euler scheme may diverge, as it happens in SDE context Hutzenthaler, Jentzen and Kloeden (2011). Therefore, one needs to seek alternative (e.g., implicit) approximations for BSDE with polynomial drivers that are also monotone and/or find conditions under which it is possible for the explicit scheme to work, as explicit schemes have certain computational advantages over implicit ones.

Our contribution.

- We extend the canonical Zhang path regularity theorem [see Ma and Zhang (2002), Imkeller and dos Reis (2010a)], originally proved under Lipschitz assumptions, to our polynomial growth monotone driver setting proving in between all the required stochastic smoothness results; essentially all first-order variations of the solution processes and estimates on the modulus of continuity.
- For our non-Lipschitz setting, we provide a thorough analysis of the family of θ-schemes, where $\theta \in [0, 1]$ characterizes the degree of implicitness of the scheme. Contrary to the FBSDEs with Lipschitz driver we show that choosing $\theta \geq 1/2$ is essential to ensure the stability of the scheme, in a similar way to the SDE context [see Mao and Szpruch (2013)]. This is to our knowledge the first result in the numerical BSDEs literature that shows a superior stability of the implicit scheme over the standard explicit one. We also generalize the concept of stability for discretization schemes [see that in Chassagneux (2012, 2013)]. This, among others things, paves a way for deriving higher order approximations schemes for FBSDEs with non-Lipschitz drivers. As an example, we prove a higher order of convergence for the trapezoidal scheme (the case $\theta = 1/2$).
- We construct an appropriately tamed version of the explicit Euler scheme for which the required stability property can be recovered. This allows us to obtain convergence of the scheme. Interestingly enough, in the special case where the driver of the FBSDEs does not depend on the SDE solution it is enough to appropriately tame the terminal condition, leaving the rest of the Euler approximation unchanged.

As a rule of thumb, implicit schemes tend to be more robust than explicit ones. Unfortunately implicit schemes involve solving an implicit equation, which creates an extra layer of complexity when compared to explicit schemes. A secondary aim of this work is to distinguish under which conditions explicit and implicit schemes can be used.

As standard in numerical analysis, we derive the global error estimates of various numerical schemes by analyzing their one-step errors and stability properties (which allows us to study how errors propagate with time). We formulate the

Fundamental Lemma [following the nomenclature from Milstein and Tretyakov (2004)] that states how to estimate the global error of a stable approximation scheme in terms of its local errors. The lemma is proved under minimal assumptions. We stress that a similar approach has been used in Chassagneux and Crisan (2012) and Chassagneux (2012, 2013); however, their results are not sufficiently general to deal with non-Lipschitz drivers.

The structure of the global error estimate given by the Fundamental Lemma allows us to study in a very easy and transparent way the special case of the θ-scheme with $\theta = 1/2$ (trapezoidal rule) which has a higher order of convergence. In this context, we also conjecture a candidate for the second-order scheme.

Concerning the implementation of the presented schemes, we propose an alternative estimator of the component Z whose standard deviation, contrary to usual estimator, does not explode as the time step vanishes.

Finally, we note that in proving convergence for the mostly-implicit schemes, we prove L^p-type uniform bounds for the scheme, thus extending the classical L^2-bound obtained previously for the discretization of Lipschitz FBSDEs [see Bouchard and Touzi (2004), Gobet and Turkedjiev (2011) and references therein].

This work is organized as follows. In Section 2, we define notation and recall standard results from the literature. In Section 3, we establish first-order variational results for the solution of the FBSDEs as well as stating the path regularity results required for the study of numerical schemes within the FBSDE framework. The remaining sections contain the discussion of several numerical schemes: in Section 4, we define the numerical discretization procedure and state general estimates for integrability and on the local errors. In Section 5, we establish the convergence of the implicit dominating schemes and in Section 6 the convergence of the tamed explicit scheme [after the terminology of Hutzenthaler, Jentzen and Kloeden (2012)]. In Section 7, we give some numerical examples.

2. Preliminaries.

2.1. *Notation.* Throughout let us fix $T > 0$. We work on a canonical Wiener space $(\Omega, \mathcal{F}, \mathbb{P})$ carrying a d-dimensional Wiener process $W = (W^1, \dots, W^d)$ restricted to the time interval $[0, T]$. We denote by $\mathcal{F} = (\mathcal{F}_t)_{t\in[0,T]}$ its natural filtration enlarged in the usual way by the $\mathbb{P}$-zero sets and by $\mathbb{E}$ and $\mathbb{E}[\cdot|\mathcal{F}_t] = \mathbb{E}_t[\cdot]$ the usual expectation and conditional expectation operator, respectively.

For vectors $x = (x^1, \dots, x^d)$ in the Euclidean space $\mathbb{R}^d$, we denote by $|\cdot|$ and $\langle\cdot,\cdot\rangle$ the canonical Euclidean norm and inner product (resp.) while $\|\cdot\|$ is the matrix norm in $\mathbb{R}^{k\times d}$ (when no ambiguity arises we use $|\cdot|$ as $\|\cdot\|$); for $A \in \mathbb{R}^{k\times d}$ A^* denotes the transpose of A; I_d denotes the d-dimensional identity matrix. For a map $b : \mathbb{R}^m \to \mathbb{R}^d$, we denote by ∇b its $\mathbb{R}^{d\times m}$-valued Jacobi matrix (gradient in case $d = 1$) whenever it exists. To denote the jth first derivative of $b(x)$ for

$x \in \mathbb{R}^m$, we write $\nabla_{x_j} b$ (valued in $\mathbb{R}^{d\times 1}$). For $b(x,y):\mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^k$, we write $\nabla_x h$ or $\nabla_y h$ to refer to its Jacobi matrix (gradient if $k=1$) with relation to x and y, respectively. Δ denotes the canonical Laplace operator.

We define the following spaces for $p>1$, $q\geq 1$, $n,m,d,k\in\mathbb{N}$: $C^{0,n}([0,T]\times\mathbb{R}^d,\mathbb{R}^k)$ is the space of continuous functions endowed with the $\|\cdot\|_\infty$-norm that are n-times continuously differentiable in the spatial variable; $C_b^{0,n}$ contains all bounded functions of $C^{0,n}$; the first superscript 0 is dropped for functions independent of time; $L^p(\mathcal{F}_t,\mathbb{R}^d)$, $t\in[0,T]$, is the space of d-dimensional $\mathcal{F}_t$-measurable RVs X with norm $\|X\|_{L^p}=\mathbb{E}[|X|^p]^{1/p}<\infty$; L^∞ refers to the subset of essentially bounded RVs; $\mathcal{S}^p([0,T]\times\mathbb{R}^d)$ is the space of d-dimensional measurable $\mathcal{F}$-adapted processes Y satisfying $\|Y\|_{\mathcal{S}^p}=\mathbb{E}[\sup_{t\in[0,T]}|Y_t|^p]^{1/p}<\infty$; $\mathcal{S}^\infty$ refers to the subset of $\mathcal{S}^p(\mathbb{R}^d)$ of absolutely uniformly bounded processes; $\mathcal{H}^p([0,T]\times\mathbb{R}^{n\times d})$ is the space of d-dimensional measurable $\mathcal{F}$-adapted processes Z satisfying $\|Z\|_{\mathcal{H}^p}=\mathbb{E}[(\int_0^T|Z_s|^2\,\mathrm{d}s)^{p/2}]^{1/p}<\infty$; $\mathbb{D}^{k,p}(\mathbb{R}^d)$ and $\mathbb{L}_{k,d}(\mathbb{R}^d)$ are the spaces of Malliavin differentiable RVs and processes; see Appendix A.2.

2.2. *Setting*. We want to study the forward–backward SDE system with dynamics (1.1)–(1.2), for $(t,x)\in[0,T]\times\mathbb{R}^d$ and $\Theta^{t,x}:=(X^{t,x},Y^{t,x},Z^{t,x})$. Here we work, for $s\in[t,T]$, with the filtration $\mathcal{F}_s^t:=\sigma(W_r-W_t:r\in[t,s])$, completed with the $\mathbb{P}$-null measure sets of $\mathcal{F}$. Concerning the functions appearing in (1.1) and (1.2) we will work with the following assumptions.

(HX0). $b:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$, $\sigma:[0,T]\times\mathbb{R}^d\to\mathbb{R}^{d\times d}$ are 1/2-Hölder continuous in their time variable, are Lipschitz continuous in their spatial variables, satisfy $\|b(\cdot,0)\|_\infty+\|\sigma(\cdot,0)\|_\infty<\infty$, and hence satisfy $|b(\cdot,x)|+|\sigma(\cdot,x)|\leq K(1+|x|)$ for some $K>0$.

(HY0). $g:\mathbb{R}^d\to\mathbb{R}^k$ is a Lipschitz function of linear growth; $f:[0,T]\times\mathbb{R}^d\times\mathbb{R}^k\times\mathbb{R}^{k\times d}\to\mathbb{R}^k$ is a continuous function and for some $L,L_x,L_y,L_z>0$ for all t,t',x,x',y,y',z,z' it holds that

$$\begin{aligned} \exists m\geq 1 \qquad & \big|f(t,x,y,z)\big|\leq L+L_x|x|+L_y|y|^m+L_z\|z\|,\\ \big\langle y'-y,f\big(t,x,y',z\big)-f(t,x,y,z)\big\rangle &\leq L_y\big|y'-y\big|^2,\\ \big|f(t,x,y,z)-f\big(t',x',y,z'\big)\big| &\leq L_t\big|t-t'\big|^{1/2}\\ &\quad +L_x\big|x-x'\big|+L_z\big\|z-z'\big\|. \end{aligned}\tag{2.1}$$

(HY0$_{\text{loc}}$). (HY0) holds and, given L_y, it holds for all t,x,y,y',z that

$$\big|f(t,x,y,z)-f\big(t,x,y',z\big)\big|\leq L_y\big(1+|y|^{m-1}+\big|y'\big|^{m-1}\big)\big|y-y'\big|.\tag{2.2}$$

(HXY1). (HX0), (HY0$_{\text{loc}}$) hold; $g\in C^1$ and $b,\sigma,f\in C^{0,1}$.

We state next a useful consequence of the monotonicity condition (2.1).

REMARK 2.1. Under assumption (HY0), for all t, x, y, y', z, z' and any $\alpha > 0$, we have

$$\begin{aligned}
&\langle y' - y, f(t, x, y', z') - f(t, x, y, z)\rangle \\
&\quad = \langle y' - y, f(t, x, y', z') \pm f(t, x, y, z') - f(t, x, y, z)\rangle \\
&\quad \le L_y|y' - y|^2 + L_z|y' - y||z' - z| \\
&\quad \le (L_y + \alpha)|y' - y|^2 + \frac{L_z^2}{4\alpha}|z' - z|^2.
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\langle y, f(t, x, y, z)\rangle \\
&\quad = \langle y - 0, f(t, x, y, z) - f(t, x, 0, z)\rangle + \langle y, f(t, x, 0, z)\rangle \\
&\quad \le L_y|y|^2 + |y|(L + L_x|x| + L_z|z|) \\
&\quad \le (L_y + \alpha)|y|^2 + \frac{3L^2}{4\alpha} + \frac{3L_x^2}{4\alpha}|x|^2 + \frac{3L_z^2}{4\alpha}|z|^2.
\end{aligned} \tag{2.3}$$

2.3. *Basic results*. In this subsection, we recall several auxiliary results concerning the solution of (1.1)–(1.2) that will become useful later. These results follows from Pardoux (1999) and Briand and Carmona (2000).

THEOREM 2.2 (Existence and uniqueness). *Let* (HX0) *and* (HY0) *hold. Then FBSDE* (1.1)–(1.2) *has a unique solution* $(X, Y, Z) \in \mathcal{S}^p \times \mathcal{S}^p \times \mathcal{H}^p$ *for any* $p \ge 2$. *Moreover, it holds for some constant* $C_p > 0$ *that*

$$\begin{aligned}
\|Y\|_{\mathcal{S}^p}^p + \|Z\|_{\mathcal{H}^p}^p &\le C_p\{\|g(X_T)\|_{L^p}^p + \|f(\cdot, X_\cdot, 0, 0)\|_{\mathcal{H}^p}^p\} \\
&\le C_p(1 + |x|^p).
\end{aligned} \tag{2.4}$$

PROOF. The existence and uniqueness results for SDE (1.1) follow from standard SDE literature. The existence and uniqueness result for the BSDE follows from Proposition 2.2 in Pardoux (1999), since the SDE results imply that $X \in \mathcal{S}^p$ for any $p \ge 2$, along with linear growth in x of g and f. The estimates for $Y \in \mathcal{S}^p$ for any $p \ge 2$ and $Z \in \mathcal{H}^p$ follow from the pathwise inequality

$$\begin{aligned}
&|Y_t|^2 + \left(1 - \frac{3L_z^2}{2\alpha}\right)\mathbb{E}_t\left[\int_t^T |Z_u|^2\,\mathrm{d}u\right] \\
&\quad \le C_{\alpha,T,t}\mathbb{E}_t\left[|g(X_T)|^2 + \int_t^T \frac{3}{4\alpha}|f(u, X_u, 0, 0)|^2\,\mathrm{d}u\right],
\end{aligned} \tag{2.5}$$

where $C_{\alpha,T,t} = \exp\{2(L_y + \alpha)(T - t)\}$, for any $\alpha > 0$ and $t \in [0, T]$. This last inequality follows from the proof of Proposition 2.2 and Exercise 2.3 in Pardoux (1999) [see also Theorem 3.6 in Briand and Carmona (2000)]. □

We now state a result concerning a priori estimates for BSDEs.

THEOREM 2.3 (A priori estimate). *Let* $p \geq 2$ *and for* $i \in \{1,2\}$, *let* $\Theta^i = (X^i, Y^i, Z^i)$ *be the solution of FBSDE* (1.1)–(1.2) *with functions* b^i, σ^i, g^i, f^i *satisfying* (HX0)–(HY0). *Then there exists* $C_p > 0$ *depending only on* p *and the constants in the assumptions such that for* $i \in \{1,2\}$

$$\begin{aligned}
&\|Y^1 - Y^2\|^p_{\mathcal{S}^p} + \|Z^1 - Z^2\|^p_{\mathcal{H}^p} \\
&\quad \leq C_p \Big\{ \mathbb{E}\Big[|g^1(X^1_T) - g^2(X^2_T)|^p \\
&\qquad + \Big(\int_0^T |f^1(s, X^1_s, Y^i_s, Z^i_s) - f^2(s, X^2_s, Y^i_s, Z^i_s)| \, \mathrm{d}s \Big)^p \Big] \Big\}.
\end{aligned} \tag{2.6}$$

PROOF. See Proposition 3.2 and Corollary 3.3 in Briand and Carmona (2000). □

COROLLARY 2.4 (Markov property and sample path continuity). *Let* (HX0) *and* (HY0) *hold. The mapping* $(t,x) \mapsto Y_t^{t,x}(\omega)$ *is continuous. There exist two* $\mathcal{B}([0,T]) \otimes \mathcal{B}(\mathbb{R}^k)$ *and* $\mathcal{B}([0,T]) \otimes \mathcal{B}(\mathbb{R}^{k\times d})$ *measurable deterministic functions* u *and* v (*resp.*) *s.t.*

$$\begin{aligned}
&Y_s^{t,x} = u(s, X_s^{t,x}), \qquad s \in [t,T], \mathrm{d}\mathbb{P}\text{-}a.s., \\
&Z_s^{t,x} = v(s, X_s^{t,x})\sigma(s, X_s^{t,x}), \qquad s \in [t,T], \mathrm{d}\mathbb{P} \times \mathrm{d}s\text{-}a.s.
\end{aligned} \tag{2.7}$$

Moreover, the Markov property holds $Y_{t+h}^{t,x} = Y_{t+h}^{t+h, X_{t+h}^{t,x}}$ *for any* $h \geq 0$ *and* $u \in C^{0,0}([0,T] \times \mathbb{R}^k)$.

PROOF. See Section 3 in Pardoux (1999). The sample path continuity of $Y_t^{t,x}$ follows from the mean-square continuity of $(Y_s^{t,x})_{s\in[t,T]}$ for $x \in \mathbb{R}^k$, $0 \leq t \leq s \leq T$, which in turn follows from inequality (2.6), combined with the Lipschitz property of $x \mapsto g(x)$ and $(t,x) \mapsto f(t,x,\cdot,\cdot)$ along with the continuity properties of $(t,x) \mapsto X_\cdot^{t,x}$ solution to (1.1).

The Markov property follows from Remark 3.1 Pardoux (1999) and the continuity of $u(t,x)$ is implied by that of $Y_t^{t,x}$. □

2.4. *Nonlinear Feynman–Kac formula.* As pointed out in the Introduction, our aim is to deepen the connection between FBSDEs and PDEs via the so-called nonlinear Feynman–Kac formula, that is, we study the probabilistic representation of the solution to a class of parabolic PDEs on $\mathbb{R}^k$ with polynomial growth coefficients that are associated with FBSDE (1.1)–(1.2). For $(t,x) \in [0,T] \times \mathbb{R}^d$, denote by $\mathcal{L}$ the infinitesimal generator of the Markov process $X^{t,x}$ solution to (1.1)

$$\mathcal{L} := \frac{1}{2} \sum_{i,j=1}^d \big([\sigma\sigma^*]_{ij}\big)(t,x)\partial^2_{x_i x_j} + \sum_{i=1}^d b_i(t,x)\partial_{x_i}, \tag{2.8}$$

and consider for a function $v = (v_1, \dots, v_k)$ the following system of backward semi-linear parabolic PDEs for $i \in \{1, \dots, k\}$: $v(T, x) = g(x)$ and

$$-\partial_t v_i(t,x) - \mathcal{L}v_i(t,x) - f_i\big(t, x, v(t,x), (\nabla v\sigma)(t,x)\big) = 0. \tag{2.9}$$

In rough, it can be easily proved using Itô's formula that if $v \in C^{1,2}([0,T] \times \mathbb{R}^d; \mathbb{R}^k)$ solves the above PDE then $Y_t := v(t, X_t)$ and $Z_t := (\nabla v\sigma)(t, X_t)$ solves BSDE (1.2) [see Proposition 3.1 in Pardoux (1999)]. But the more interesting result is the converse one, that is, that $u(t,x) := Y_t^{t,x}$ is the solution of the PDE (in some sense). It was established in Theorem 3.2 of Pardoux (1999) (recalled next) that indeed $(t,x) \mapsto Y_t^{t,x}$ is the viscosity solution of the PDE.

THEOREM 2.5. *Let* (HX0), (HY0) *hold and take* $(t,x) \in [0,T] \times \mathbb{R}^d$. *Furthermore, assume that the* i*th component of the driver function* f *depends only on the* i*th row of the matrix* $z \in \mathbb{R}^{k\times d}$, *that is*, $f_i(t,x,y,z) = f_i(t,x,y,z^i)$.

Then $u(t,x) := Y_t^{t,x}$ *is a continuous function of* (t,x) *that grows at most polynomially at infinity and is a viscosity solution of* (2.9) [*in the sense of Definition* 3.2 *in Pardoux* (1999)].

REMARK 2.6 (Multi-dimensional case). The proof of Theorem 2.5 relies on a BSDE comparison theorem that holds only in the case $k = 1$ (i.e., when Y is one-dimensional). Nonetheless, with the restriction imposed by (HY0), it is still possible to use the said comparison theorem to prove Theorem 2.5, we point the reader to Theorem 2.4 and Remark 2.5 in Pardoux (1999).

It is possible to show that $(t,x) \mapsto Y_t^{t,x}$ is the solution to (2.9) not only in the viscosity sense, but also in weak sense (in weighted Sobolev spaces), this has been done in Matoussi and Xu (2008) and Zhang and Zhao (2012).

2.5. *Examples*. One equation covered by our setting is the FitzHugh–Nagumo PDE with recovery, used in biology and related to the modeling of the electrical distribution of the heart or the potential in neurons.

EXAMPLE 2.7 (The FH–N equation with recovery). Let $(t,x) \in [0,T] \times \mathbb{R}^d$, $g = (g_u, g_v)$, $f = (f_u, f_v)$ and $g, f, (u,v) : [0,T] \times \mathbb{R}^d \to \mathbb{R}^2$. The FH–N PDE has the dynamics: $u(T,\cdot) = g_u(\cdot)$, $v(T,\cdot) = g_v(\cdot)$ and

$$-\partial_t u - \tfrac{1}{2}\Delta u - f_u(u,v) = 0, \qquad -\partial_t v - \Delta v - f_v(u,v) = 0,$$

where $f_u(u,v) = u - u^3 + v$ and $f_v(u,v) = u - v$. f clearly satisfies (HY0) and ($\mathrm{HY0}_{\mathrm{loc}}$).

A simpler setup of the above model is its one-dimensional version.

EXAMPLE 2.8 (FH–N equation without recovery). For $(t,x) \in [0,T] \times \mathbb{R}$ the FH–N equation without recovery is described by

$$-\partial_t u - \tfrac{1}{2}\Delta u - \big(cu^3 + bu^2 - au\big) = 0, \qquad u(T,x) = g(x). \tag{2.10}$$

When $c = -1$, $b = 1 + a$, $a \in \mathbb{R}$ and with the choice of $g(x) = (1 + e^x)^{-1}$, one can verify that the C_b^∞ solution u to (2.10) is given by

$$u(t,x) = \big(1 + \exp\{x - (1/2 - a)(T - t)\}\big)^{-1} \in C_b^\infty\big([0,T] \times \mathbb{R}\big). \tag{2.11}$$

The FBSDE corresponding to this PDE is given by (1.1)–(1.2) with the following data:

$$b(\cdot,\cdot) = 0; \qquad \sigma(\cdot,\cdot) = 1; \qquad f(t,x,y,z) = cy^3 + by^2 - ay;$$
$$c = -1; \qquad b = 1 + a,$$

and the terminal condition function g is given above. Both (HX0) and (HY0$_{\text{loc}}$) hold (for any a, notice that $u \ge 0$ for any a) and the theory we develop throughout applies to this class of examples. We will use the case $a = -1$ in our simulations.

3. Representation results, path regularity and other properties. As seen before $u(t,x) := Y_t^{t,x}$ is a viscosity solution of PDE (2.9). If $u \in C^{1,2}$, we would also obtain the representation of the process Z as $Z_t^{t,x} = (\nabla_x u \sigma)(t,x)$, but in view of Theorem 2.5 we have not given meaning to $\nabla_x u$. The main aim of this section is to first prove some representation formulas, that express Z as a function of Y and X, then use these representation formulas to obtain the so-called L^2- (and L^p-) path regularity results needed to prove the convergence of the numerical discretization of FBSDE (1.1)–(1.2) in the later sections. A by-product of these results is the existence of $\nabla_x u$.

3.1. *Differentiability in the spatial parameter.* Take the system (1.1)–(1.2) into account. We now show that the smoothness of the FBSDE parameters b, σ, g, f carries over to the solution process $\Theta = (X, Y, Z)$.

THEOREM 3.1. *Let* (HXY1) *hold and* $(t,x) \in [0,T] \times \mathbb{R}^d$.

Then u [from (2.7)] is continuously differentiable in its spatial variable. Moreover, the triple $\nabla_x \Theta^{t,x} = (\nabla_x X^{t,x}, \nabla_x Y^{t,x}, \nabla_x Z^{t,x}) \in \mathcal{S}^p \times \mathcal{S}^p \times \mathcal{H}^p$ *for any* $p \ge 2$ *and solves for* $0 \le t \le s \le T$

$$\begin{cases} \nabla_x X_s^{t,x} = I_d + \displaystyle\int_t^s (\nabla_x b)\big(r, X_r^{t,x}\big) \nabla_x X_r^{t,x} \,\mathrm{d}r \\ \qquad\qquad + \displaystyle\int_t^s (\nabla_x \sigma)\big(r, X_r^{t,x}\big) \nabla_x X_r^{t,x} \,\mathrm{d}W_r, \\ \nabla_{x_i} Y_s^{t,x} = (\nabla_x g)\big(X_T^{t,x}\big) \nabla_{x_i} X_T^{t,x} - \displaystyle\int_s^T \nabla_{x_i} Z_r^{t,x} \,\mathrm{d}W_r \\ \qquad\qquad + \displaystyle\int_t^T F\big(r, \nabla_{x_i} \Theta_r^{t,x}\big) \,\mathrm{d}r \end{cases} \tag{3.1}$$

for $i \in \{1, \ldots, d\}$ *and with*[5]

$$F : (\omega, r, x, \chi, \Upsilon, \Gamma) \mapsto (\nabla_x f)\big(r, \Theta_r^{t,x}\big) \cdot \chi + (\nabla_y f)\big(r, \Theta_r^{t,x}\big) \cdot \Upsilon + (\nabla_z f)\big(r, \Theta_r^{t,x}\big) \cdot \Gamma.$$

There exists a positive constant C_p *independent of* x *such that*

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \big\|\big(\nabla_x Y^{t,x}, \nabla_x Z^{t,x}\big)\big\|_{\mathcal{S}^p\times\mathcal{H}^p} \le C_p. \tag{3.2}$$

Furthermore, *for* u *as in* (2.7) *we have for* $x \in \mathbb{R}^d$ *and* $0 \le t \le s \le T$

$$\begin{aligned} &\nabla_x Y_s^{t,x} = (\nabla_x u)\big(s, X_s^{t,x}\big)\nabla_x X_s^{t,x}, \qquad \mathbb{P}\text{-a.s.} \quad \text{and} \\ &\|\nabla_x u\|_\infty < \infty. \end{aligned} \tag{3.3}$$

We recall that $\nabla_x Y^{t,x}$ is $\mathbb{R}^{k\times d}$-valued and $\nabla_{x_i} Y^{t,x}$ denotes its ith column we use a similar notation follows for $\nabla_x X$ and $\nabla_x Z$.

PROOF OF THEOREM 3.1. Throughout fix $(t,x) \in [0,T] \times \mathbb{R}^d$ and let $\{e_i\}_{i\in\{1,\ldots,d\}}$ be the canonical unit vectors of $\mathbb{R}^d$. Let $i \in \{1, \ldots, d\}$.

The results concerning SDE (1.1) follow from those in Section 2.5 in Imkeller and dos Reis (2010a). We start by showing that the partial derivatives $(\nabla_{x_i} Y^{t,x}, \nabla_{x_i} Z^{t,x})$ for any i exist, then we will show the full differentiability. We start by proving that (3.1) has indeed a solution for every i. Unfortunately, the driver of (3.1) does not satisfy (HY0), and hence we cannot quote Theorem 2.2 directly; we use a more general result from Briand et al. (2003). We remark though, that the techniques used to obtain moment estimates of the form of (2.4) and (2.6) are the same in both Briand et al. (2003) and Pardoux (1999).

FBSDE (3.1) has a unique solution $\Xi^{t,x,i} := (\nabla_{x_i} X^{t,x}, U^{t,x,i}, V^{t,x,i}) \in \mathcal{S}^p \times \mathcal{S}^p \times \mathcal{H}^p$ for any $p \ge 2$, where (U^i, V^i) replaces $(\nabla_{x_i} Y, \nabla_{x_i} Z)$. This follows by a direct application of Theorem 4.2 in Briand et al. (2003). It is easy to see that under (HXY1) the conditions (H1)–(H5) in Briand et al. [(2003), pages 118–119] are satisfied. First, under (HXY1), standard SDE theory [see, e.g., Theorem 2.4 in Imkeller and dos Reis (2010a)] ensures that $\nabla_x X \in \mathcal{S}^p$ for all $p \ge 2$, which along with $\nabla_x g, \nabla_x f \in C_b^{0,0}$, implies in turn that the terminal condition $(\nabla_x g)(X_T^{t,x})\nabla_{x_i} X_T^{t,x} \in L^p_{\mathcal{F}_T}$ and the term $(\nabla_x f)(\cdot, \Theta_\cdot^{t,x})\nabla_{x_i} X_\cdot^{t,x} = F(\cdot, \nabla_{x_i} X_\cdot^{t,x}, 0, 0) \in \mathcal{S}^p$ for any $p \ge 2$. Given the linearity of F and the Lipschitz property of f in its z-variable, it follows that F is uniformly Lipschitz in Γ. More-

[5]The term $(\nabla_z f)(\cdot, \Theta) \cdot \Gamma$ can be better understood if one interprets z in f not as in $\mathbb{R}^{k\times d}$ but as $(\mathbb{R}^d)^k$, that is, f receives not a matrix but its $\mathbb{R}^d$-valued k lines.

over, since f satisfies (2.1) it implies that F is monotone[6] in Υ, that is,

$$\big\langle \Upsilon - \Upsilon', (\nabla_y f)\big(\cdot, \Theta^{t,x}_{\cdot}\big) \cdot (\Upsilon - \Upsilon')\big\rangle \le L_y |\Upsilon - \Upsilon'|^2 \qquad \forall \Upsilon, \Upsilon' \in \mathbb{R}^k. \tag{3.4}$$

The continuity of $\Upsilon \mapsto F(r, x, \chi, \Upsilon, \Gamma)$ is also clear. Finally, the linearity of F, the fact that $\Theta \in \mathcal{S}^p \times \mathcal{S}^p \times \mathcal{H}^p$ for any $p \ge 2$ and (2.2) implies that condition (H5) in Briand et al. (2003) is also satisfied, that is, that for any $R > 0$, $\sup_{|\Upsilon| \le R} |F(r, x, \nabla_{x_i} X^{t,x}_r, \Upsilon, 0) - F(r, x, \nabla_{x_i} X^{t,x}_r, 0, 0)| \in L^1([t, T] \times \Omega)$. We are therefore under the conditions of Theorem 4.2 in Briand et al. (2003), as claimed.

In view of (2.3) and the linearity of F one can obtain moment estimates in the style of (2.4) by following arguments similar to those in the proof of Theorem 2.2 [recall that (2.3) takes in this case a very simple form]. In view of (2.4), we have (recall that $\nabla X \in \mathcal{S}^p$ for all $p \ge 2$)

$$\begin{aligned} &\|U^i\|^p_{\mathcal{S}^p} + \|V^i\|^p_{\mathcal{H}^p} \\ &\quad \le C_p \big\{ \big\|(\nabla_x g)\big(X^{t,x}_T\big) \nabla_{x_i} X^{t,x}_T\big\|^p_{L^p} + \big\|(\nabla_x f)\big(\cdot, \Theta^{t,x}_{\cdot}\big) \nabla_{x_i} X^{t,x}_{\cdot}\big\|^p_{\mathcal{H}^p}\big\} \\ &\quad \le C_p \|\nabla_{x_i} X^{t,x}\|^p_{\mathcal{S}^p} \le C_p, \end{aligned} \tag{3.5}$$

where C_p does *not* depend on x, t or i.

In order to obtain results on the first-order variation of the solution, we follow standard BSDE techniques used already in Imkeller and dos Reis (2010a), Briand and Confortola (2008) or dos Reis, Réveillac and Zhang (2011); we start by studying the behavior of $\Theta^{t,x+\varepsilon e_i} - \Theta^{t,x}$ for any $\varepsilon > 0$. Take $h \in \mathbb{R}^d$. Via the stability of SDEs and inequality (2.6) [and (HY0)], it is clear that a constant $C_p > 0$ independent of x exists such that

$$\begin{aligned} \lim_{h \to 0} \|\Theta^{t,x+h} - \Theta^{t,x}\|_{\mathcal{S}^p \times \mathcal{S}^p \times \mathcal{H}^p} &\le \lim_{h \to 0} C_p \|X^{x+h} - X^x\|_{\mathcal{S}^p} \\ &\le \lim_{h \to 0} C_p |h| = 0. \end{aligned} \tag{3.6}$$

Define

$$\begin{aligned} \delta\Theta^{\varepsilon,i} &:= \big(\delta X^{\varepsilon,i}, \delta Y^{\varepsilon,i}, \delta Z^{\varepsilon,i}\big) \\ &:= \big(\Theta^{t,x+\varepsilon e_i} - \Theta^{t,x}\big)/\varepsilon - \big(\nabla_{x_i} X^{t,x}, U^{t,x,i}, V^{t,x,i}\big) \end{aligned}$$

for which

$$\begin{aligned} \delta Y^{\varepsilon,i}_s = {}& \bigg[\frac{1}{\varepsilon}\big(g\big(X^{t,x+\varepsilon e_i}_T\big) - g\big(X^{t,x}_T\big)\big) - (\nabla_x g)\big(X^{t,x}_T\big) \nabla_{x_i} X^{t,x}_T\bigg] - \int_s^T \delta Z^{\varepsilon,i}_r \,\mathrm{d}W_r \\ &+ \int_s^T \bigg[\frac{1}{\varepsilon}\big(f\big(r, \Theta^{t,x+\varepsilon e_i}_r\big) - f\big(r, \Theta^{t,x}_r\big)\big) \\ &\qquad - F\big(r, x, \nabla_{x_i} X^{t,x}_r, U^{t,x,i}_r, V^{t,x,i}_r\big)\bigg] \,\mathrm{d}r. \end{aligned} \tag{3.7}$$

[6] This follows easily from the differentiability of f, its monotonicity in y and the definition of directional derivative.

Using the differentiability of the involved functions, we can re-write (3.7) as a linear FBSDE with random coefficients satisfying in its essence a (HY0) type assumption: for $s \in [t, T]$, $j \in \{1, \dots, d\}$

$$\begin{cases} \delta X_s^{\varepsilon,j} = 0 + \int_t^s \big[b_x^{\varepsilon,j}(r)\delta X_r^{\varepsilon,j} + \delta\nabla b_r^\varepsilon \nabla_{x_j} X_r^{t,x}\big]\,\mathrm{d}r \\ \qquad\qquad + \int_t^s \big[\sigma_x^{\varepsilon,j}(r)\delta X_r^{\varepsilon,j} + \delta\nabla\sigma_r^\varepsilon \nabla_{x_j} X_r^{t,x}\big]\,\mathrm{d}W_r, \\ \delta Y_s^{\varepsilon,i} = \big[g_x^{\varepsilon,i}(T)\delta X_T^{\varepsilon,i} + \delta\nabla g_T^\varepsilon \nabla_{x_i} X_T^{t,x}\big] - \int_s^T \delta Z_r^{\varepsilon,i}\,\mathrm{d}W_r \\ \qquad\qquad + \int_s^T \big[f_x^{\varepsilon,i}(r)\delta X_r^{\varepsilon,i} + f_y^{\varepsilon,i}(r)\delta Y_r^{\varepsilon,i} + f_z^{\varepsilon,i}(r)\delta Z_r^{\varepsilon,i} \\ \qquad\qquad\qquad + \delta\nabla f_r^\varepsilon \cdot \big(\nabla_{x_i} X_r^{t,x}, U_r^{t,x,i}, V_r^{t,x,i}\big)\big]\,\mathrm{d}r, \end{cases} \tag{3.8}$$

where $\delta\nabla f$ and $\delta\nabla\varphi$ denote the differences

$$\delta\nabla f_\cdot^\varepsilon := \big(f_x^{\varepsilon,i}, f_y^{\varepsilon,i}, f_z^{\varepsilon,i}\big)(\cdot) - (\nabla_x f, \nabla_y f, \nabla_z f)\big(\cdot, \Theta_\cdot^{t,x}\big)$$

and

$$\delta\nabla\varphi_\cdot^\varepsilon := \varphi_x^{\varepsilon,i}(\cdot) - \nabla_x\varphi\big(\cdot, \Theta_\cdot^{t,x}\big),$$

for $\varphi \in \{b, \sigma, g\}$ (with some abuse of notation) and $r \in [t, T]$, and where we defined

$$\begin{aligned} \varphi_x^{\varepsilon,i}(r) &:= \int_0^1 (\nabla_x\varphi)\big(r, (1-\lambda)X_r^{t,x} + \lambda X_r^{t,x+\varepsilon e_i}\big)\,\mathrm{d}\lambda \\ &= \int_0^1 (\nabla_x\varphi)\big(r, X_r^{t,x} + \lambda\big(X_r^{t,x+\varepsilon e_i} - X_r^{t,x}\big)\big)\,\mathrm{d}\lambda, \end{aligned}$$

and $f_*^{\varepsilon,i}$ for $* \in \{x, y, z\}$ in the following way:

$$\begin{aligned} f_z^{\varepsilon,i}(r) &:= \int_0^1 (\nabla_z f)\big(r, X_r^{t,x+\varepsilon e_i}, Y_r^{t,x+\varepsilon e_i}, Z_r^{t,x} + \lambda\big(Z_r^{t,x+\varepsilon e_i} - Z_r^{t,x}\big)\big)\,\mathrm{d}\lambda, \\ f_y^{\varepsilon,i}(r) &:= \int_0^1 (\nabla_y f)\big(r, X_r^{t,x+\varepsilon e_i}, Y_r^{t,x} + \lambda\big(Y_r^{t,x+\varepsilon e_i} - Y_r^{t,x}\big), Z_r^{t,x}\big)\,\mathrm{d}\lambda, \\ f_x^{\varepsilon,i}(r) &:= \int_0^1 (\nabla_x f)\big(r, X_r^{t,x} + \lambda\big(X_r^{t,x+\varepsilon e_i} - X_r^{t,x}\big), Y_r^{t,x}, Z_r^{t,x}\big)\,\mathrm{d}\lambda. \end{aligned}$$

The assumptions imply immediately that $b_x^{\varepsilon,i}, \sigma_x^{\varepsilon,i}, f_x^{\varepsilon,i}, f_z^{\varepsilon,i}$ are uniformly bounded, while $f_y^{\varepsilon,i} \in \mathcal{S}^p$, $p \geq 2$ (thanks to $\mathrm{HY0_{loc}}$). Furthermore, using estimate (2.4) [along with $\|X^{t,x}\|_{\mathcal{S}^p}^p \leq C_p(1 + |x|^p)$], (3.5), (3.6), the continuity of $\varphi \in \{b, \sigma, g\}$ and its derivative it is easy to see that, in combination with the dominated convergence theorem, one has

$$\begin{aligned} \lim_{\varepsilon \to 0} \big\{ &\big\|\varphi_x^{\varepsilon,i}(\cdot) - \nabla_x\varphi\big(\cdot, \Theta_\cdot^{t,x}\big)\big\|_{\mathcal{S}^p} \\ &+ \big\|\big(f_x^{\varepsilon,i}, f_y^{\varepsilon,i}, f_z^{\varepsilon,i}\big)(\cdot) - (\nabla_x f, \nabla_y f, \nabla_z f)\big(\cdot, \Theta_\cdot^{t,x}\big)\big\|_{\mathcal{H}^p}\big\} = 0. \end{aligned} \tag{3.9}$$

We remark that in the above limit a localization argument for the convergence of $f_y^{\varepsilon,i}(\cdot)$ to $\nabla_y f(\cdot,\Theta_\cdot)$ is required, namely that we work inside a ball (of any given radius) centered around x in which all points $x+\varepsilon e_i \in \mathbb{R}^d$ as ε vanishes are contained. We do not detail the argumentation since it is similar to that given in, for example, Imkeller and dos Reis (2010a), Briand and Confortola (2008) or dos Reis, Réveillac and Zhang (2011).

With this in mind we return to (3.7), written in the form of (3.8), and since it is a linear FBSDE satisfying the monotonicity condition (2.1) we have via Corollary 3.3 in Briand and Carmona (2000) [essentially our moment estimate (2.4) for FBSDE (3.8)] in combination with (3.5), (3.6) and (3.9), that for any i

$$\lim_{\varepsilon\to 0}\left\|\frac{1}{\varepsilon}\big(\Theta^{t,x+\varepsilon e_i}-\Theta^{t,x}\big)-\big(\nabla_{x_i}X^{t,x},U^{t,x,i},V^{t,x,i}\big)\right\|_{\mathcal{S}^p\times\mathcal{S}^p\times\mathcal{H}^p}=0 \qquad \forall p\ge 2.$$

Since the limit exists we identify $(\nabla_{x_i}Y^{t,x},\nabla_{x_i}Z^{t,x})$ with $(U^{t,x,i},V^{t,x,i})$ and, moreover, estimate (3.5) implies estimate (3.2). Furthermore, the above limit implies in particular that (take $s=t$)

$$\begin{aligned}\nabla_{x_i}u(t,x)&=\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\big[u(t,x+\varepsilon e_i)-u(t,x)\big]\\&=\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\big[Y_t^{t,x+\varepsilon e_i}-Y_t^{t,x}\big]=\nabla_{x_i}Y_t^{t,x}.\end{aligned}$$

Observing that the RHS of (3.5) is a constant independent of $t\in[0,T]$, $x\in\mathbb{R}^d$ and $i\in\{1,\dots,d\}$ we can conclude that

$$\|\nabla_{x_i}u\|_\infty=\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\big|\nabla_{x_i}Y_t^{t,x}\big|<\infty. \tag{3.10}$$

It is clear that $(\nabla_{x_i}Y_s^{t,x})_{s\in[t,T]}$ is continuous in its time parameter as it is a solution to a BSDE; we now focus on the continuity of $x\mapsto\nabla_{x_i}Y_t^{t,x}$. Let $x,x'\in\mathbb{R}^d$. The difference $\nabla_{x_i}Y^{t,x}-\nabla_{x_i}Y^{t,x'}$ is the solution to a *linear* FBSDE following from (3.1). As before, it is easy to adapt the computations and apply Corollary 3.3 in Briand and Carmona (2000) [essentially our moment estimate (2.6) for FBSDEs (3.1)] to the difference $\nabla_{x_i}Y_s^{t,x}-\nabla_{x_i}Y_s^{t,x'}$ yielding

$$\begin{aligned}&\|\nabla_{x_i}Y^{t,x}-\nabla_{x_i}Y^{t,x'}\|_{\mathcal{S}^2}^2\\&\quad\le C_p\bigg\{\big\|(\nabla_x g)\big(X_T^{t,x}\big)\nabla_{x_i}X_T^{t,x}-(\nabla_x g)\big(X_T^{t,x'}\big)\nabla_{x_i}X_T^{t,x'}\big\|_{L^2}^2\\&\qquad+\mathbb{E}\bigg[\bigg(\int_0^T\big|F\big(r,x,\nabla_{x_i}X_r^{t,x},\nabla_{x_i}Y_r^{t,x},\nabla_{x_i}Z_r^{t,x}\big)\\&\qquad\qquad-F\big(r,x',\nabla_{x_i}X_r^{t,x'},\nabla_{x_i}Y_r^{t,x},\nabla_{x_i}Z_r^{t,x}\big)\big|\,\mathrm{d}s\bigg)^p\bigg]\bigg\}.\end{aligned}$$

Given the known results on SDEs, the linearity of F, (3.5), the continuity of the derivatives of f and (3.6), dominated convergence theorem yields that $\|\nabla_{x_i} Y^{t,x} - \nabla_{x_i} Y^{t,x'}\|^2_{\mathcal{S}^2} \to 0$ as $x' \to x$ uniformly on compact sets. This mean-square continuity of $\nabla_{x_i} Y^{t,x}$ implies in particular that $\nabla_{x_i} Y^{t,x}_t = \nabla_{x_i} u(t,x)$ is continuous. In conclusion, we just proved that for any $i \in \{1,\dots,d\}$ the partial derivatives $\nabla_{x_i} u$ exist and are continuous; hence, standard multi-dimensional real analysis implies that u is continuously differentiable in its spatial variables. This argumentation is similar to that in the proof of Corollary 2.4.

We are left to prove (3.3). Note that for any $\varepsilon > 0$ we have $(Y^{t,x+\varepsilon e_i}_s - Y^{t,x}_s)/\varepsilon = (u(s, X^{t,x+\varepsilon e_i}_s) - u(s, X^{t,x}_s))/\varepsilon$. By sending $\varepsilon \to 0$ and using the (continuous) differentiability of u, we have $\nabla_x Y^{t,x}_s = (\nabla_x u)(s, X^{t,x}_s)\nabla_x X^{t,x}_s$. Hence, as the RHS of (3.5) is a constant independent of $t \in [0,T]$, $x \in \mathbb{R}^d$ and i we can conclude (let $s \searrow t$) that $\|\nabla_x u\|_\infty = \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |\nabla_x Y^{t,x}_t| < \infty$. □

3.2. *Malliavin differentiability.* As in the previous section, we show a form of regularity of the solution Θ to (1.1)–(1.2), namely the stochastic variation of Θ in the sense of Malliavin's calculus.

THEOREM 3.2 (Malliavin differentiability). *Let* (HXY1) *hold. Then the solution* $\Theta = (X, Y, Z)$ *of* (1.1)–(1.2) *verifies*:

- $X \in \mathbb{L}^{1,2}$ *and* DX *admits a version* $(u,t) \mapsto D_u X_t$ *satisfying for* $0 \le u \le t \le T$

$$D_u X_t = \sigma(u, X_u) + \int_u^t (\nabla_x b)(s, X_s) D_u X_s \, \mathrm{d}s + \int_u^t (\nabla_x \sigma)(s, X_s) D_u X_s \, \mathrm{d}W_s.$$

Moreover, for any $p \ge 2$ *there exists* $C_p > 0$ *such that*

$$\sup_{u\in[0,T]} \|D_u X\|^p_{\mathcal{S}^p} \le C_p\bigl(1 + |x|^p\bigr). \tag{3.11}$$

- *For any* $0 \le t \le T$, $x \in \mathbb{R}^m$ *we have* $(Y, Z) \in \mathbb{L}^{1,2} \times (\mathbb{L}^{1,2})^d$. *A version of* $(DY, DZ)_{0\le u,t\le T}$ *satisfies*: *for* $t < u \le T$, $D_u Y_t = 0$ *and* $D_u Z_t = 0$, *and for* $0 \le u \le t$,

$$\begin{aligned} D_u Y_t = (\nabla_x g)(X_T) D_u X_T &+ \int_t^T \bigl\langle (\nabla f)(s, \Theta_s), D_u \Theta_s \bigr\rangle \, \mathrm{d}s \\ &- \int_t^T D_u Z_s \, \mathrm{d}W_s. \end{aligned} \tag{3.12}$$

Moreover, $(D_t Y_t)_{0\le t\le T}$ *defined by the above equation is a version of* $(Z_t)_{0\le t\le T}$.
- *The following representation holds for any* $0 \le u \le t \le T$ *and* $x \in \mathbb{R}^m$:

$$D_u X_t = \nabla_x X_t (\nabla_x X_u)^{-1} \sigma(u, X_u) \mathbb{1}_{[0,u]}(t), \tag{3.13}$$

$$D_u Y_t = \nabla_x Y_t (\nabla_x X_u)^{-1} \sigma(u, X_u), \qquad \textit{a.s.}, \tag{3.14}$$

$$Z_t = \nabla_x Y_t (\nabla_x X_t)^{-1} \sigma(s, X_t), \qquad \textit{a.s.} \tag{3.15}$$

REMARK 3.3 (Y is already in $\mathbb{L}^{1,2}$). Via Theorem 3.1, we know that $u \in C^{0,1}$. Under (HXY1) it is known that $X \in \mathbb{L}^{1,2}$ [see Nualart (2006)], hence using the chain rule [for Malliavin calculus, see Proposition 1.2.3 in Nualart (2006)] we obtain $Y_\cdot = u(\cdot, X_\cdot) \in \mathbb{L}_{1,2}$. A careful analysis of Theorem 3.1 and the results about $\nabla_x u$ show that indeed $X, Y \in \mathbb{L}^{1,p}$ for all $p \geq 2$ [just combine (3.11) with (A.1) as described in Appendix A.2].

Using the fact that $X, Y \in \mathbb{L}^{1,2}$, the statement of Theorem 3.2 follows easily if the driver f in (1.2) does not depend on z. One would argue in the following way: for any $t \in [0, T]$

$$\left(g(X_T) - Y_t + \int_t^T f(r, X_r, Y_r)\,\mathrm{d}r\right)_{t\in[0,T]} \in \mathbb{L}^{1,2}$$
$$\Rightarrow \left(\int_t^T Z_r\,\mathrm{d}W_r\right)_{t\in[0,T]} \in \mathbb{L}^{1,2} \Leftrightarrow Z \in \mathbb{L}^{1,2},$$

this follows from the definition of the BSDE (1.2) itself and Theorem A.3. The dynamics of (3.12) and the representation formulas (3.14), (3.15) follow by arguments similar to those given below.

PROOF OF THEOREM 3.2. The first part of the statement is trivial as it follows from standard SDE theory; see, for example, Nualart (2006) or Theorem 2.5 in Imkeller and dos Reis (2010a). To prove the other statements of the theorem, we will use an identification trick by taking advantage of the fact we already know that $Y \in \mathbb{L}^{1,2}$ (see Remark 3.3).

Let (X, Y, Z) be the solution of (1.1)–(1.2) and define the following BSDE:

$$U_t = g(X_T) + \int_t^T \widehat{f}(r, V_r)\,\mathrm{d}r - \int_t^T V_r\,\mathrm{d}W_r, \tag{3.16}$$

where the driver $\widehat{f} : \Omega \times [0, T] \times \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\widehat{f}(t, v) := f(t, X_t, Y_t, v) = f\big(t, X_t, u(t, X_t), v\big). \tag{3.17}$$

It is clear that: $g(X_T) \in \mathbb{D}^{1,2}$, $f(\cdot, X_\cdot, Y_\cdot, 0) \in \mathbb{L}^{1,p}$ for all $p \geq 2$ (see Remark 3.3) and that $v \mapsto \widehat{f}(\cdot, v)$ is a Lipschitz continuous function, all these imply in particular via Lipschitz BSDE theory [see Theorem 2.1, Proposition 2.1 in El Karoui, Peng and Quenez (1997)] that there exists a pair $(U, V) \in \mathcal{S}^2 \times \mathcal{H}^2$ solving (3.16). Furthermore, Theorem 2.2 in El Karoui, Peng and Quenez (1997) states that the solution to (1.2) is unique, and hence the solution of (3.16) verifies $(U, V) = (Y, Z)$.

Proposition 5.3 in El Karoui, Peng and Quenez (1997), yields the existence of the Malliavin derivatives (DU, DV) of (U, V) with the following dynamics. Set $\Xi := (X, Y, V)$, then for $t < u \leq T$ we have $D_u U_t = 0$, $D_u V_t = 0$ and for $0 \le u \le t$

$$D_u U_t = (\nabla_x g)(X_T) D_u X_T + \int_t^T \langle (\nabla f)(s, \Xi_s), (D_u \Xi_s)\rangle \,\mathrm{d}s - \int_t^T D_u V_s \,\mathrm{d}W_s.$$

Since $(U, V) = (Y, Z)$ then from the above BSDE for (DU, DV) follows BSDE (3.12). Moreover, Proposition 5.9 in El Karoui, Peng and Quenez (1997) yields (3.14) and (3.15) for (U, V) which carry out for (Y, Z). □

3.3. *Representation results.* Here, we combine the results of the two previous subsections to obtain representation formulas that will allow us to establish the path regularity properties of Y and Z required for the convergence proof of the numerical discretization.

THEOREM 3.4. *Let* (HXY1) *hold, then the following representation holds*:

$$Z_s^{t,x} = (\nabla_x u\sigma)\big(s, X_s^{t,x}\big), \qquad 0 \le t \le s \le T, \mathrm{d}\mathbb{P}\text{-}a.s., \tag{3.18}$$

$$= \nabla_x Y_s^{t,x}\big(\nabla_x X_s^{t,x}\big)^{-1}\sigma\big(s, X_s^{t,x}\big), \qquad 0 \le t \le s \le T, \mathrm{d}\mathbb{P}\text{-}a.s., \tag{3.19}$$

and $\|Z\|_{\mathcal{S}^q}^q \le C_q(1 + |x|^q)$, $q \ge 2$.

Assume that only (HX0) *and* ($\mathrm{HY0_{loc}}$) *hold, then for some* $C > 0$ *it holds* $|Z_t| \le C|\sigma(X_t)|$ $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-*a.s. and in particular*

$$|Z_t| \le C\big(1 + |X_t|\big), \qquad \mathrm{d}t \otimes \mathrm{d}\mathbb{P}\text{-}a.s. \tag{3.20}$$

PROOF. We first prove all the results under (HXY1), then argue via mollification that (3.20) holds under (HX0)–($\mathrm{HY0_{loc}}$).

Proof under (HXY1). The representation $Z = \nabla Y(\nabla X)^{-1}\sigma(\cdot, X)$ follows from Theorem 3.2, while from Theorem 3.1, we have

$$\begin{aligned} Z_s^{t,x} &= \nabla_x Y_s^{t,x}\big(\nabla_x X_s^{t,x}\big)^{-1}\sigma\big(s, X_s^{t,x}\big) \\ &= (\nabla_x u)\big(s, X_s^{t,x}\big)\big(\nabla_x X_s^{t,x}\big(\nabla_x X_s^{t,x}\big)^{-1}\big)\sigma\big(s, X_s^{t,x}\big) \\ &= (\nabla_x u)\big(s, X_s^{t,x}\big)\sigma\big(s, X_s^{t,x}\big). \end{aligned}$$

Since all the involved processes (in the RHS) are continuous, we can identify Z with its continuous version. Moreover, as all the processes in the RHS belong to $\mathcal{S}^p$ for all $p \ge 2$ it follows that $Z \in \mathcal{S}^p$ for all $p \ge 2$. Combining Hölder's inequality with the fact that $X, \nabla X \in \mathcal{S}^p$ for all $p \ge 2$ and estimate (3.2), leads to (3.20), that is,

$$\begin{aligned} \|Z\|_{\mathcal{S}^p} &= \big\|\nabla_x Y_\cdot^{t,x}\big(\nabla_x X_\cdot^{t,x}\big)^{-1}\sigma\big(\cdot, X_\cdot^{t,x}\big)\big\|_{\mathcal{S}^p} \\ &\le C_p\big\|\nabla_x Y^{t,x}\big\|_{\mathcal{S}^{3p}}\big\|(\nabla_x X)^{-1}\big\|_{\mathcal{S}^{3p}}\big\|1 + X^{t,x}\big\|_{\mathcal{S}^{3p}} \\ &\le C_p\big(1 + |x|\big). \end{aligned} \tag{3.21}$$

A careful inspection of the used inequalities shows that the constant C_p in (3.21) depends only on the several constants appearing in the assumptions (HX0)–(HY0$_{\text{loc}}$).

Proof of (3.20) *under* (HX0)–(HY0$_{\text{loc}}$). In this step, we rely on a standard mollification arguments similar to those in the proof of Theorem 5.2 in Imkeller and dos Reis (2010a). Note that a driver satisfying (HY0$_{\text{loc}}$) once mollified will still satisfy assumption (HY0$_{\text{loc}}$) with the same constants.

Take b^n, σ^n, g^n, f^n as mollified versions of b, σ, g, f in their spatial variables such that the mollified functions satisfy uniformly (in n) (HX0) and (HY0$_{\text{loc}}$), with uniform Lipschitz and monotonicity constants. Theorem 2.2 ensures that $\Theta = (X^n, Y^n, Z^n) \in \mathcal{S}^p \times \mathcal{S}^p \times \mathcal{H}^p$ for any $p \geq 2$ and solves (1.1)–(1.2) with b^n, σ^n, g^n, f^n replacing b, σ, g, f. Since the mollified functions satisfy (HXY1), it follows from the above proof that for each fixed n we have $Z^n \in \mathcal{S}^p$. Moreover, in view of (2.6) and the standard theory of SDEs it is rather simple to deduce that $\Theta^n \to \Theta$ as $n \to \infty$ in $\mathcal{S}^p \times \mathcal{S}^p \times \mathcal{H}^p$ for all $p \geq 2$. Let u^n denote the solution to the PDE linked to FBSDE (1.1)–(1.2) with data b^n, σ^n, g^n, f^n and we drop the superscript (t, x) and work with (X^n, Y^n, Z^n).

From (3.18), we have $|Z_s^n| = |(\nabla_x u^n \sigma^n)(s, X_s^n)|$ at least $\mathrm{d}s \otimes \mathrm{d}\mathbb{P}$-a.s. From (3.10) [or (3.2)], we can conclude that $|\nabla_x Y_t^{t,x,n}| = |\nabla_x u^n(t, x)| \leq C$, with C independent of n, and hence quite easily that

$$|Z_s^n| \leq C|\sigma^n(s, X_s^n)| \leq C(1 + |X_s^n|), \qquad \mathrm{d}s \otimes \mathrm{d}\mathbb{P}\text{-a.s.}, \tag{3.22}$$

where we last used the linear growth condition of σ^n.

Finally combine: the pointwise convergence of $\sigma^n \to \sigma$ (knowing that all σ^n and σ have the same Lipschitz constant); the fact that $X^n \to X$ in $\mathcal{S}^p$ (standard SDE stability theory); and Theorem 2.3 yielding that $Z^n \to Z$ in $\mathcal{H}^p$ to conclude that (3.22) holds in the limit. □

3.4. *Path regularity results.* Now let π be a partition of the interval $[0, T]$, say $0 = t_0 < \cdots < t_i < \cdots < T_N = T$, and mesh size $|\pi| = \max_{i=0,\dots,N-1}(t_{i+1} - t_i)$. Given π, define $r_\pi = |\pi|/(\min_{i=0,\dots,N-1}(t_{i+1} - t_i))$.

Let Z be the control process in the solution to BSDE (1.2), under (HX0)–(HY0). We define a set of random variables $\{\bar{Z}_{t_i}\}_{t_i \in \pi}$ termwise given by

$$\begin{aligned} \bar{Z}_{t_i} &= \frac{1}{t_{i+1} - t_i} \mathbb{E}\left[\int_{t_i}^{t_{i+1}} Z_s \,\mathrm{d}s \Big| \mathcal{F}_{t_i}\right], \qquad 0 \leq i \leq N-1 \quad \text{and} \\ \bar{Z}_{t_N} &= Z_T. \end{aligned} \tag{3.23}$$

The RV Z_T can be obtained using (3.18), namely $Z_T = (\nabla_x g)(X_T)\sigma(T, X_T)$ when $g \in C^1$. If g is only Lipschitz continuous then one easily sees that a RV $G \in L^\infty(\mathcal{F}_T)$ exists such that $Z_T = G\sigma(T, X_T)$. In any case, under (HX0) and

(HY0) it easily follows that

$$\bar{Z}_{t_N} = Z_T \in L^p(\mathcal{F}_T) \qquad \text{for any } p \geq 2 \quad \text{and} \tag{3.24}$$
$$\bar{Z}_{t_i} \in L^2 \qquad \text{for any } t_i \in \pi.$$

It is not difficult to show that $\bar{Z}_{t_i}$ is the best $\mathcal{F}_{t_i}$-measurable square integrable RV approximating Z in $\mathcal{H}^2([t_i, t_{i+1}])$, that is,

$$\mathbb{E}\left[\int_{t_i}^{t_{i+1}} |Z_s - \bar{Z}_{t_i}|^2 \,\mathrm{d}s\right] = \inf_{\xi \in L^2(\Omega, \mathcal{F}_{t_i})} \mathbb{E}\left[\int_{t_i}^{t_{i+1}} |Z_s - \xi|^2 \,\mathrm{d}s\right]. \tag{3.25}$$

Let now $\bar{Z}_t := \bar{Z}_{t_i}$ for $t \in [t_i, t_{i+1})$, $0 \leq i \leq N-1$. It is equally easy to see that $\bar{Z}$ converges to Z in $\mathcal{H}^2$ as $|\pi|$ vanishes: since Z is adapted, the family of processes Z^π indexed by our partition defined by $Z^\pi_t = Z_{t_i}$ for $t \in [t_i, t_{i+1})$ converges to Z in $\mathcal{H}^2$ as $|\pi|$ goes to zero. Since $\{\bar{Z}\}$ is the best $\mathcal{H}^2$-approximation of Z, we obtain

$$\|Z - \bar{Z}\|_{\mathcal{H}^2} \leq \|Z - Z^\pi\|_{\mathcal{H}^2} \to 0 \qquad \text{as } |\pi| \to 0,$$

although without knowing the rate of this convergence.

The next result expresses the modulus of continuity (in the time variable) for Y and Z.

THEOREM 3.5 (Path regularity). *Let* (HX0), (HY0$_{\text{loc}}$) *hold. Then the unique solution* (X, Y, Z) *to* (1.1)–(1.2) *satisfies* $(X, Y, Z) \in \mathcal{S}^p \times \mathcal{S}^p \times \mathcal{H}^p$ *for all* $p \geq 2$. *Moreover*:

(i) *for any* $p \geq 2$ *there exists a constant* $C_p > 0$ *such that for* $0 \leq s \leq t \leq T$ *we have*

$$\mathbb{E}\Big[\sup_{s \leq u \leq t} |Y_u - Y_s|^p\Big] \leq C_p\big(1 + |x|^p\big)|t - s|^{p/2}; \tag{3.26}$$

(ii) *for any* $p \geq 2$ *there exists a constant* $C_p > 0$ *such that for any partition* π *of* $[0, T]$ *with mesh size* $|\pi|$

$$\sum_{i=0}^{N-1} \mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} |Z_t - Z_{t_i}|^2 \,\mathrm{d}t\right)^{p/2} + \left(\int_{t_i}^{t_{i+1}} |Z_t - Z_{t_{i+1}}|^2 \,\mathrm{d}t\right)^{p/2}\right] \tag{3.27}$$
$$\leq C_p\big(1 + |x|^p\big)|\pi|^{p/2};$$

(iii) *in particular, there exists a constant* C *such that for any partition* $\pi = \{0 = t_0 < \cdots < t_N = T\}$ *of the interval* $[0, T]$ *with mesh size* $|\pi|$ *we have*

$$\mathrm{REG}_\pi(Y)^2 := \max_{0 \leq i \leq N-1} \sup_{t \in [t_i, t_{i+1}]} \{\mathbb{E}[|Y_t - Y_{t_i}|^2] + \mathbb{E}[|Y_t - Y_{t_{i+1}}|^2]\}$$
$$\leq C|\pi|$$

and $\sum_{i=0}^{N-1}\mathbb{E}[\int_{t_i}^{t_{i+1}}|Z_s-\bar{Z}_{t_i}|^2\,\mathrm{d}s]\le C|\pi|$. *Moreover, if* r_π *remains bounded*[7] *as* $|\pi|\to 0$ *then*

$$\begin{aligned}\mathrm{REG}_\pi(Z)^2 := \sum_{i=0}^{N-1}\mathbb{E}\Big[\int_{t_i}^{t_{i+1}}|Z_s-\bar{Z}_{t_i}|^2\,\mathrm{d}s\Big]\\ + \sum_{i=0}^{N-1}\mathbb{E}\Big[\int_{t_i}^{t_{i+1}}|Z_s-\bar{Z}_{t_{i+1}}|^2\,\mathrm{d}s\Big]\\ \le C|\pi|.\end{aligned}$$

PROOF. Fix $(t,x)\in[0,T]\times\mathbb{R}^d$, take $s\in[t,T]$ and throughout this proof we work with $\Theta^{t,x}$ and $\nabla_x\Theta^{t,x}$; to avoid a notational overload we omit the super- and subscript and write Θ and $\nabla\Theta$. Under the theorem's assumptions, $(X,Y,Z)\in\mathcal{S}^p\times\mathcal{S}^p\times\mathcal{H}^p$ for all $p\ge 2$ and (3.20) holds. We first prove points (i) and (ii) under assumption (HXY1), then we use the same mollification argument as in the proof of (3.20) to recover the case (HX0)–($\mathrm{HY0_{loc}}$). We then explain how (iii) is obtained.

Proof of (i) *under* (HXY1). From Theorem 3.4 follows $Z\in\mathcal{S}^q$ for any $q\ge 2$. Writing the BSDE for the difference Y_u-Y_s for $0\le s\le u\le T$, we have

$$\begin{aligned}Y_u-Y_s &= \int_s^u f(r,\Theta_r)\,\mathrm{d}r-\int_s^u Z_r\,\mathrm{d}W_r\\ &\le \int_s^u K\big(1+|X_r|+|Y_r|^m+|Z_r|\big)\,\mathrm{d}r-\int_s^u Z_r\,\mathrm{d}W_r.\end{aligned}$$

Taking absolute values, the sup over $u\in[s,t]\subseteq[0,T]$, power p, expectations and Jensen's inequality leads, for some constant $C_p>0$, to

$$\begin{aligned}&\mathbb{E}\Big[\sup_{u\in[s,t]}|Y_u-Y_s|^p\Big]\\ &\quad\le C_p\Big\{|t-s|^p\big(1+\|(X,Y,Z)\|^p_{\mathcal{S}^p\times\mathcal{S}^p\times\mathcal{S}^p}\big)+\mathbb{E}\Big[\sup_{u\in[s,t]}\Big|\int_s^u Z_r\,\mathrm{d}W_r\Big|^p\Big]\Big\}.\end{aligned}$$

Applying BDG to the last term in the RHS, then (3.20) yields

$$\begin{aligned}&\mathbb{E}\Big[\sup_{u\in[s,t]}\Big|\int_s^u Z_r\,\mathrm{d}W_r\Big|^p\Big]\\ &\quad\le C_p\mathbb{E}\Big[\Big(\int_s^t|Z_r|^2\,\mathrm{d}r\Big)^{p/2}\Big]\\ &\quad\le C_p\mathbb{E}\Big[\Big(\int_s^t|1+X_r|^2\,\mathrm{d}r\Big)^{p/2}\Big]\le C_p|t-s|^{p/2}\|X\|^p_{\mathcal{S}^p}.\end{aligned}$$

[7] This is trivially satisfied for the uniform grid for which $r_\pi=1$.

It then follows that

$$\mathbb{E}\Big[\sup_{u\in[s,t]}|Y_u - Y_s|^p\Big] \le C_p\{|t-s|^p + |t-s|^{p/2}\} \le C_p(1+|x|^p)|t-s|^{p/2}.$$

Proof of (ii) *under* (HXY1). To prove the desired inequality, we use the representation (3.15) [alternatively (3.19)]. We first estimate the difference $\mathbb{E}[(\int_{t_i}^{t_{i+1}}|Z_s - Z_{t_i}|^2\,\mathrm{d}s)^{p/2}]$. The difference $Z_s - Z_{t_i}$ can be written as $Z_s - Z_{t_i} = I_1 + I_2$ with $I_2 := (\nabla Y_s - \nabla Y_{t_i})(\nabla X_{t_i})^{-1}\sigma(t_i, X_{t_i})$ and

$$I_1 := \nabla Y_s\{((\nabla X_s)^{-1} - (\nabla X_{t_i})^{-1})\sigma(s, X_s) + (\nabla X_{t_i})^{-1}[\sigma(s, X_s) - \sigma(t_i, X_{t_i})]\}.$$

The estimation of I_1 is rather easy as it relies on Hölder's inequality combined with (3.2), (HX0), Theorems 2.3 and 2.4 in Imkeller and dos Reis (2010a) [see proof of Theorem 5.5(i) in Imkeller and dos Reis (2010a)], in short we have

$$\mathbb{E}[|I_1|^p] \le C_p(1+|x|^p)|\pi|^{p/2}.$$

Concerning the second part, the estimation of I_2, it follows from an adaptation of the proof of Theorem 5.5(ii) in Imkeller and dos Reis (2010b). We reformulate the main argument and skip the obvious details. Let us start with a simple trick, as $s \in [t_i, t_{i+1}]$,

$$\begin{aligned}\mathbb{E}\big[&|(\nabla Y_s - \nabla Y_{t_i})(\nabla X_{t_i})^{-1}\sigma(t_i, X_{t_i})|^p\big] \\ &= \mathbb{E}\big[\mathbb{E}[|\nabla Y_s - \nabla Y_{t_i}|^p|\mathcal{F}_{t_i}]|(\nabla X_{t_i})^{-1}\sigma(t_i, X_{t_i})|^p\big].\end{aligned} \tag{3.28}$$

Writing the BSDE for the difference $\nabla Y_s - \nabla Y_{t_i}$ for $t_i \le s \le t_{i+1}$, we have for some constant $C > 0$

$$\mathbb{E}[|\nabla Y_s - \nabla Y_{t_i}|^p|\mathcal{F}_{t_i}] \le C\mathbb{E}[\widehat{I}_{[t_i,t_{i+1}]}|\mathcal{F}_{t_i}],$$

where

$$\widehat{I}_{[t_i,t_{i+1}]} := \left(\int_{t_i}^{t_{i+1}}|(\nabla f)(r,\Theta_r)||\nabla\Theta_r|\,\mathrm{d}r\right)^p + \left(\int_{t_i}^{t_{i+1}}|\nabla Z_r|^2\,\mathrm{d}r\right)^{p/2},$$

where we used the conditional BDG inequality and maximized over the time interval $[t_i, t_{i+1}]$.

Combining these last two inequalities and observing that since ∇X_{t_i} and $\sigma(X_{t_i})$ are $\mathcal{F}_{t_i}$-adapted, we can drop the conditional expectation from (3.28). Hence, for some $C > 0$,

$$\begin{aligned}&\sum_{i=0}^{N-1}\mathbb{E}\Big[\Big(\int_{t_i}^{t_{i+1}}|I_2|^2\,\mathrm{d}s\Big)^{p/2}\Big] \\ &\qquad \le C|\pi|^{p/2-1}\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\mathbb{E}[|I_2|^p]\,\mathrm{d}s\end{aligned}$$

$$
\begin{aligned}
&\le C|\pi|^{p/2-1}\sum_{i=0}^{N-1}|\pi|\mathbb{E}\big[\big|(\nabla X_{t_i})^{-1}\sigma(t_i,X_{t_i})\big|^p\widehat{I}_{[t_i,t_{i+1}]}\big]\\
&\le C|\pi|^{p/2}\mathbb{E}\bigg[\sup_{0\le t\le T}\big|(\nabla X_t)^{-1}\sigma(t,X_t)\big|^p\sum_{i=0}^{N-1}\widehat{I}_{[t_i,t_{i+1}]}\bigg]\\
&\le C|\pi|^{p/2}\big\|(\nabla X)^{-1}\big\|_{\mathcal{S}^{3p}}^{1/3}\|1+X\|_{\mathcal{S}^{3p}}^{1/3}\|\widehat{I}_{[0,T]}\|_{L^1}\\
&\le C\big(1+|x|^p\big)|\pi|^{p/2}.
\end{aligned}
$$

The last line follows from standard inequalities (sum of powers is less than the power of the sum), the growth conditions on ∇f and the fact that for any $q\ge 2$ we have: X, ∇X, $(\nabla X)^{-1}\in\mathcal{S}^q$, Y, $\nabla Y\in\mathcal{S}^q$, (3.20) and $\nabla Z\in\mathcal{H}^q$.

Collecting now the estimates, we obtain the desired result for the difference $Z_s-Z_{t_i}$. To have the same estimate for the difference $Z_s-Z_{t_{i+1}}$ we need only to repeat the above calculations with a minor change in order to incorporate the $Z_{t_{i+1}}$: one writes $Z_s-Z_{t_{i+1}}$ with the help of I_1^{i+1} and I_2^{i+1}, which are I_1 and I_2, respectively, but with t_{i+1} instead of t_i. The estimate for I_1^{i+1} follows from SDE theory in the same fashion as for I_1 above; concerning I_2^{i+1} one just needs another small trick,

$$
\begin{aligned}
I_2^{i+1}&=(\nabla Y_s-\nabla Y_{t_{i+1}})(\nabla X_{t_{i+1}})^{-1}\sigma(t_{i+1},X_{t_{i+1}})\\
&\le\big(|\nabla Y_s|+|\nabla Y_{t_{i+1}}|\big)\big[(\nabla X_{t_{i+1}})^{-1}\sigma(t_{i+1},X_{t_{i+1}})-(\nabla X_{t_i})^{-1}\sigma(t_i,X_{t_i})\big] \qquad (3.29)\\
&\quad+(\nabla Y_s-\nabla Y_{t_{i+1}})(\nabla X_{t_i})^{-1}\sigma(t_i,X_{t_i}). \qquad (3.30)
\end{aligned}
$$

The rest of the proof follows just like before, like I_1 for (3.29) and like I_2 for (3.30).

Final step—(i) *and* (ii) *under* (HX0)–($\mathrm{HY0_{loc}}$)—*arguing via mollification*: Here, we follow the same setup as in the proof of (3.20) under (HX0)–($\mathrm{HY0_{loc}}$) (see Theorem 3.4).

Take b^n,σ^n,g^n,f^n as mollified versions of b,σ,g,f in their spatial variables such that the mollified functions satisfy uniformly (in n) (HX0) and ($\mathrm{HY0_{loc}}$), with uniform Lipschitz and monotonicity constant. From the proof of Theorem 3.4, we know that $\Theta=(X^n,Y^n,Z^n)\in\mathcal{S}^p\times\mathcal{S}^p\times\mathcal{H}^p$ for any $p\ge 2$ and $\Theta^n\to\Theta$ as $n\to\infty$ in $\mathcal{S}^p\times\mathcal{S}^p\times\mathcal{H}^p$ for all $p\ge 2$.

For each $n\in\mathbb{N}$ estimates (3.26) and (3.27) hold for Θ^n. Since b^n,σ^n,g^n,f^n satisfy (HX0) and ($\mathrm{HY0_{loc}}$) uniformly in n then it is easy to check that the constants appearing on the RHS of (3.26) and (3.27) are independent of n. Hence, by taking the limit of $n\to\infty$ in (3.26) and (3.27) and given the convergence $\Theta^n\to\Theta$ as $n\to\infty$ (and the continuity of the involved functions) the statement follows.

Proof of (iii) *under* (HX0)–($\mathrm{HY0_{loc}}$). The estimates concerning Y and $\bar{Z}_{t_i}$ follow trivially from (3.26) on the one hand, and (3.27) combined with (3.25) on the other hand. For the difference $Z_s - \bar{Z}_{t_{i+1}}$, more care is required,

$$\begin{aligned}
&\sum_{i=0}^{N-1} \mathbb{E}\Big[\int_{t_i}^{t_{i+1}} |Z_s - \bar{Z}_{t_{i+1}}|^2 \, \mathrm{d}s\Big] \\
&\quad \le 2 \sum_{i=0}^{N-1} \mathbb{E}\Big[\int_{t_i}^{t_{i+1}} |Z_s - Z_{t_{i+1}}|^2 + |Z_{t_{i+1}} - \bar{Z}_{t_{i+1}}|^2 \, \mathrm{d}s\Big] \\
&\quad \le C|\pi| + 2 \sum_{i=0}^{N-1} (t_{i+1} - t_i) \mathbb{E}\big[|Z_{t_{i+1}} - \bar{Z}_{t_{i+1}}|^2\big],
\end{aligned}$$

where the last inequality follows from the proof of (ii). We next estimate the last term in the RHS, since $\bar{Z}_{t_N} = Z_T$ by construction

$$\begin{aligned}
&\sum_{i=0}^{N-1} (t_{i+1} - t_i) \mathbb{E}\big[|Z_{t_{i+1}} - \bar{Z}_{t_{i+1}}|^2\big] \\
&\quad = \sum_{i=0}^{N-2} (t_{i+1} - t_i) \mathbb{E}\big[|Z_{t_{i+1}} - \bar{Z}_{t_{i+1}}|^2\big] \\
&\quad \le r_\pi \sum_{i=0}^{N-2} (t_{i+2} - t_{i+1}) \mathbb{E}\big[|Z_{t_{i+1}} - \bar{Z}_{t_{i+1}}|^2\big] \\
&\quad \le r_\pi \sum_{i=0}^{N-2} \int_{t_{i+1}}^{t_{i+2}} \mathbb{E}\big[|Z_{t_{i+1}} - \bar{Z}_{t_{i+1}}|^2\big] \, \mathrm{d}s \\
&\quad \le r_\pi \sum_{j=1}^{N-1} \int_{t_j}^{t_{j+1}} \mathbb{E}\big[|Z_{t_j} - \bar{Z}_{t_j}|^2\big] \, \mathrm{d}s \\
&\quad \le 2r_\pi \sum_{i=0}^{N-1} \mathbb{E}\Big[\int_{t_i}^{t_{i+1}} |Z_s - Z_{t_i}|^2 + |Z_s - \bar{Z}_{t_i}|^2 \, \mathrm{d}s\Big],
\end{aligned}$$

where we made use of the assumption on the grid. The result now follows by combining (iii) with the above estimates and having in mind that r_π is uniform over the partition. □

COROLLARY 3.6. *Let* (HX0), (HY0) *hold and take the family* $\{\bar{Z}_{t_i}\}_{t_i \in \pi}$. *For any* $p \ge 1$ *there exists constant* C_p *independent of* $|\pi|$ *such that*

$$\mathbb{E}\Bigg[\sum_{i=0}^{N-1} \big(|\bar{Z}_{t_i}|^2 (t_{i+1} - t_i)\big)^p\Bigg] \le C_p < \infty.$$

If, moreover, ($\mathrm{HY0}_{\mathrm{loc}}$) *holds then* $\max_{t_i \in \pi} \mathbb{E}[|\bar{Z}_{t_i}|^{2p}] \le C_p < \infty$.

PROOF. The second statement follows easily from the definition of $\bar{Z}_{t_i}$ [see (3.23)] and the fact that estimate (3.20) holds under (HY0$_{\text{loc}}$). Moreover, under this assumption the second estimate implies the first.

We leave the proof of the first statement for the interested reader. The proof is based on standard integral manipulations combining the definition of $\bar{Z}$, Jensen's inequality, the fact that $Z \in \mathcal{H}^p$ and the tower property of the conditional expectation [see Section 4.7.5 in Lionnet (2014)]. □

3.5. *Some finer properties*. Here, we discuss properties of the solution to (1.1)–(1.2) in more specific settings. The first lemma concerns a set-up where Z belongs to $\mathcal{S}^\infty$ (rather than $\mathcal{H}^2$ or $\mathcal{S}^2$).

PROPOSITION 3.7 (The additive noise case). *Let* (HX0)–(HY0$_{\text{loc}}$) *hold. Assume additionally that* $\sigma(t,x) = \sigma(t)$ *for all* $(t,x) \in [0,T] \times \mathbb{R}^d$. *Then* $Z \in \mathcal{S}^\infty$.

PROOF. Assume first that (HXY1) also hold. Then the result follows easily by combining the representation formula (3.18) with the 2nd part of (3.3) and injecting that σ is uniformly bounded.

Now using a standard mollification argument, as was used in the last step of the proof of Theorem 3.5, one easily concludes that the result also holds under (HX0)–(HY0$_{\text{loc}}$). □

If the initial data g and $f(\cdot,\cdot,0,0)$ are bounded, then so will be the Y process; the second component, Z will also satisfy a type of boundedness condition [see (3.31) below].

LEMMA 3.8 (The bounded setting). *Let* (HX0), (HY0) *hold and further that* g *and* $(t,x) \mapsto f(t,x,0,0)$ *are uniformly bounded then* $(Y,Z) \in \mathcal{S}^\infty \times \mathcal{H}^2$.

Denoting $\mathcal{T}_{[0,T]}$ *the set of all stopping times* $\tau \in [0,T]$, *then* Z *satisfies further*[8] *for some constant* $K_{\text{BMO}} > 0$

$$\sup_{\tau \in \mathcal{T}_{[0,T]}} \left\| \mathbb{E}\left[\int_\tau^T |Z_s|^2 \,\mathrm{d}s \Big| \mathcal{F}_\tau \right] \right\|_\infty \le K_{\text{BMO}} < \infty. \tag{3.31}$$

The constant K_{BMO} *depends only on* $\|Y\|_{\mathcal{S}^\infty}$, *the bounds for* g, $f(\cdot,\cdot,0,0)$ *and the constants appearing in* (HY0).

PROOF. The boundedness of Y follows from (2.5) by using that $g(X_\cdot)$ and $f(\cdot, X_\cdot, 0, 0)$ are in $\mathcal{S}^\infty$. Knowing that $Y \in \mathcal{S}^\infty$ we can easily adapt the proof of Lemma 10.2 in Touzi (2013) to our setting, where we make use of the inequality $|z| \le 1 + |z|^2$, to obtain (3.31); an alternative proof would be to use (2.5). □

[8]This means Z belongs to the so-called $\mathcal{H}_{\text{BMO}}$-spaces, see Section 2.3 in Imkeller and dos Reis (2010a) or Section 10.1 in Touzi (2013).

The first of the above results implies that Z is bounded. Such a setting also includes the case of $\sigma(t,x)=1$ which is common in many applications in reaction–diffusion equations. The next result provides another type of control for the growth of the process Z without the boundedness assumption on σ.

PROPOSITION 3.9. *Let the assumptions of Lemma 3.8 hold. Assume further that $|Z|^2$ is a submartingale then $|Z_t| \le K_{\mathrm{BMO}}/\sqrt{T-t}$, $\forall t \in [0,T]$ $\mathbb{P}$-a.s.*

In particular, if σ is uniformly elliptic and (HXY1) *holds then there exists $C>0$ such that $|\nabla_x u(t,x)| \le C/\sqrt{T-t}$, $\forall (t,x) \in [0,T) \times \mathbb{R}^n$.*

PROOF. The first statement follows by a careful but rather clean analysis of the fact that Z satisfies (3.31), which in particular means any $t \in [0,T]$ $\mathbb{P}$-a.s.

$$K_{\mathrm{BMO}} \ge \mathbb{E}\Big[\int_t^T |Z_s|^2 \,\mathrm{d}s \Big| \mathcal{F}_t\Big] = \int_t^T \mathbb{E}\big[|Z_s|^2|\mathcal{F}_t\big]\,\mathrm{d}s$$
$$\ge \int_t^T |Z_t|^2 \,\mathrm{d}s = |Z_t|^2(T-t),$$

where we applied Fubini then used the submartingale property of Z^2. The sought statement now follows by a direct rewriting of the above inequality. The second statement in the proposition follows from the first by using the representation $Z_t^{t,x} = (\nabla_x u\sigma)(t,x)$ and the ellipticity of σ. □

4. Numerical discretization and general estimates. In this section and the following ones, we discuss the numerical approximation of (1.1)–(1.2). We consider a regular partition[9] π of $[0,T]$ with $N+1$ points $t_i = ih$ for $i = 0,\dots,N$ with $h := T/N$.

REMARK 4.1 (On constants). Throughout the rest of this work, we introduce a generic constant $c>0$, that will always be independent of h or N, though it may depend on the problem's data, namely the constants appearing in the assumptions, and may change from line to line.

4.1. *Discretization of the SDE and further setup.* Numerical methods for SDEs with Lipschitz continuous coefficients are well understood; see Section 10 in Kloeden and Platen (1992). Therefore, we take as given a family of random variables $\{X_i\}_{i=0,\dots,N}$ that approximates the solution X to (1.1) over the grid π. More exactly, for any $p \ge 2$ there exists a constant $c = c(T,p,x)$ such that

$$\sup_{N\in\mathbb{N}} \max_{i=0,\dots,N} \mathbb{E}\big[|X_i|^p\big] \le c \tag{4.1}$$

[9]We point out that the results we state would hold for nonuniform time-steps, but we work with a regular partition for notational clarity and to keep the focus on the main issues.

and

$$\mathrm{ERR}_{\pi,p}(X) := \max_{i=0,\dots,N} \mathbb{E}\big[|X_{t_i} - X_i|^p\big]^{1/p} \le ch^\gamma, \qquad \gamma \ge \frac{1}{2}, \tag{4.2}$$

where γ is called the rate of the strong convergence and the random variables $\{X_{t_i}\}_{t_i\in\pi}$ are the solution to (1.1) on the grid points π. Under (HX0), the Euler scheme give an approximation with $\gamma = 1/2$. For conditions required for the higher order schemes, we refer to Kloeden and Platen (1992). Since the upper bound in the estimate on the error on X does not depend on p, and since we use only the case $p = 2$ in the following, we simplify the notation to $\mathrm{ERR}_\pi(X) \le ch^\gamma$.

Throughout the rest of this work, we assume that the family $\{X_i\}_{i=0,\dots,N}$ has been computed; we denote by $\{\mathcal{F}_i\}_{i=0,\dots,N}$ the associated discrete-time filtration $\mathcal{F}_i := \sigma(X_j, j = 0, \dots, i)$ and with respect to this filtration we define the operator $\mathbb{E}_i[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_i]$.

For the analysis of the time-discretization error, we also make use of the following standard path-regularity estimate for X, which holds under (HX0): there exists a constant $c > 0$ such that

$$\begin{aligned}\mathrm{REG}_\pi(X) := &\max_{i=0,\dots,N-1} \sup_{t_i\le s\le t_{i+1}} \big\{\mathbb{E}\big[|X_s - X_{t_i}|^2\big]^{1/2} + \mathbb{E}\big[|X_s - X_{t_{i+1}}|^2\big]^{1/2}\big\}\\ &\le ch^{1/2}.\end{aligned} \tag{4.3}$$

4.2. *Schemes considered and main convergence results*. For the reader's convenience, we state immediately the numerical schemes under consideration as well as their convergence rates. The rest of this work deals with the proofs of the stated results.

Theorem 3.5 implies that to approximate (Y, Z) solution to (1.2) over $[0, T]$ one needs only to approximate the family $\{(Y_{t_i}, \bar{Z}_{t_i})\}_{t_i\in\pi}$ [recall (3.23)] on the grid π via a family of random variables $\{(Y_i, Z_i)\}_{i=0,\dots,N}$, the said numerical approximation. The error criterion we consider is given by

$$\mathrm{ERR}_\pi(Y, Z) := \left(\max_{i=0,\dots,N} \mathbb{E}\big[|Y_{t_i} - Y_i|^2\big] + \sum_{i=0}^{N-1} \mathbb{E}\big[|\bar{Z}_{t_i} - Z_i|^2\big]h\right)^{1/2}. \tag{4.4}$$

4.2.1. *The implicit-dominant θ-schemes of Section 5*. Let $\theta \in [0, 1]$. Define $Y_N := g(X_N)$ and $Z_N := 0$ and, for $i = N-1, N-2, \dots, 0$,

$$\begin{aligned}Y_i := &\mathbb{E}_i\big[Y_{i+1} + (1-\theta)f(t_{i+1}, X_{i+1}, Y_{i+1}, Z_{i+1})h\big]\\ &+ \theta f(t_i, X_i, Y_i, Z_i)h,\end{aligned} \tag{4.5}$$

$$Z_i := \mathbb{E}_i\left[\frac{\Delta W_{i+1}}{h}\big(Y_{i+1} + (1-\theta)f(t_{i+1}, X_{i+1}, Y_{i+1}, Z_{i+1})h\big)\right], \tag{4.6}$$

where $\Delta W_{i+1} = W_{t_{i+1}} - W_i$. The above scheme is the called θ-scheme. Its derivation is presented in Section 4.4 and the solvability (in Y_i) of (4.5) for $\theta > 0$ is discussed in Section 4.5. When $\theta = 1$ this is the implicit backward Euler scheme, when $\theta = 0$ this is the explicit scheme. For $\theta \in]0, 1[$ it is a combination of both. The particular case of $\theta = 1/2$ is the trapezoidal scheme which, we will show, has a better convergence rate (under certain conditions). The convergence rate of the above scheme is summarized in the next result.

THEOREM 4.2. *Let* (HX0), (HY0_{loc}) *hold as well as the restriction* $h \le \min\{1, [4\theta(L_y + 3d\theta L_z^2)]^{-1}\}$. *Let* $\gamma \ge 1/2$ *be the order of the approximation* $\{X_i\}_{i=0,\dots,N}$ *of* X *as in* (4.1). *Then, for the scheme* (4.5)–(4.6) *we have*:

(i) *For* $\theta \in [1/2, 1]$, *there exists a constant* c *such that* $\text{ERR}_\pi(Y, Z) \le ch^{1/2}$.

(ii) *Take* $\theta = 1/2$ *and scheme* (4.5). *Assume that* $f \in C^2$, $f(t, x, y, z) = f(y)$ *and* $\partial_{yy}^2 f$ *has at most polynomial growth*, *then there exists* $c > 0$ *such that* $\max_{i=0,\dots,N} \mathbb{E}[|Y_{t_i} - Y_i|^2]^{1/2} \le ch^{\min\{7/4,\gamma\}}$.

Reasons why the above theorem only holds for $\theta \ge 1/2$—that is to say when the scheme is "more implicit than explicit"—will be seen later in the proofs in Section 5. But from the motivating example of the Introduction, we know already that one could not have expected convergence of the scheme in general, for all $\theta \in [0, 1]$.

4.2.2. *The tamed explicit scheme of Section* 6. By inspecting the proof of Lemma A.2, we see that the unboundedness of $g(X_T)$ plays the key role in the explosion. In Section 6, we analyze a tamed version of the fully explicit ($\theta = 0$) scheme (4.5)–(4.6).

For any level $L > 0$, we define the truncation function $T_L : \mathbb{R} \to \mathbb{R}$, $x \mapsto -L \vee x \wedge L$. We denote similarly its extension as a function from $\mathbb{R}^d$ to $\mathbb{R}^d$ (projection on the ball of radius L). We consider the following scheme: define $Y_N := T_{L_h}(g(X_N))$, $Z_N := 0$, and for $i = N - 1, \dots, 0$,

$$Y_i := \mathbb{E}_i\big[Y_{i+1} + f\big(t_{i+1}, T_{K_h}(X_{i+1}), Y_{i+1}, Z_{i+1}\big)h\big], \tag{4.7}$$

$$Z_i := \mathbb{E}_i\bigg[\frac{\Delta W_{i+1}}{h}\big(Y_{i+1} + f\big(t_{i+1}, T_{K_h}(X_{i+1}), Y_{i+1}, Z_{i+1}\big)h\big)\bigg], \tag{4.8}$$

where the levels L_h and K_h satisfy $e^{c_1 T}(L_h^2 + c_2 T + c_2 T K_h^2) \le h^{-1/(m-1)}$, with

$$c_1 = 2\big(L_y + 12dL_z^2 + 2L_y^2\big) \quad \text{and} \quad c_2 = \max\bigg\{\frac{L^2}{4dL_z^2}, \frac{L_x^2}{4dL_z^2}\bigg\}.$$

For $h \le h^*$, where h^* satisfies $e^{c_1 T} c_2 T \le (h^*)^{-1/(m-1)}/3$ and $h^* \le 1/(32dL_z^2)$ we can take

$$L_h = \frac{1}{\sqrt{3}} e^{-(1/2)c_1 T}\bigg(\frac{1}{h}\bigg)^{1/(2(m-1))} \quad \text{and} \quad K_h = \frac{1}{\sqrt{3}} \frac{e^{-(1/2)c_1 T}}{\sqrt{c_2 T}}\bigg(\frac{1}{h}\bigg)^{1/(2(m-1))}.$$

Concerning the scheme (4.7)–(4.8), we have the following convergence rate.

THEOREM 4.3. *Let* (HX0), ($\mathrm{HY0_{loc}}$) *hold and* $h \le h^*$. *Assume that the order* γ *of the approximation* $\{X_i\}_{i=0,\dots,N}$ *of* X *is at least* $1/2$ [*see* (4.1)]. *Then for the controlled explicit scheme* (4.7)–(4.8), *there exists a constant c such that* $\mathrm{ERR}_\pi(Y, Z) \le ch^{1/2}$.

4.2.3. *Modus operandi for the proofs and organization of rest of the paper.* The proof of the above results is a (long) two-step procedure. The first step is contained in the rest of this section since it is a general argument common to most discretization schemes. The second one is scheme-specific, hence the separation into Sections 5 and 6. We now describe the said procedure.

Before one is able to state a global error estimate for (4.4), one needs to find the local error estimates, that is, the distance between the solution and its approximation over one time interval $[t_i, t_{i+1}]$. This local error has two components. The first is the *one-step discretization error* following from approximating the involved integrals over $[t_i, t_{i+1}]$ by some quadrature rule. The second is the backward propagation of the error due to not having at time t_{i+1} the true solution to compute the approximation at time t_i and we coin it *stability error*.

In the next subsection, we give the *Fundamental Lemma for convergence* (Lemma 4.6) that explains how to aggregate the one-step discretization error and the stability error for each $[t_i, t_{i+1}]$ into a single estimate with (4.4) on its LHS. This later allows us to derive the convergence rates.

The estimation of the one-step discretization error is common to both schemes. This is done in Section 4.6 and the general result is stated in Proposition 4.13. Left to Sections 5 and 6 is the scheme-specific stability analysis [i.e., the estimation of $\mathcal{R}^{\mathcal{S}}(H)$ in (4.11) below]. Sections 5 and 6 follow the same structure: (1) one first shows some uniform global integrability for the scheme; (2) then one studies the local (one-step) stability of the scheme; this shows how the error propagates in just one backward step, and yields an expression for the terms H_j composing the stability remainder (see Definition 4.4 below); (3) one finally estimates the stability remainder $\mathcal{R}^{\mathcal{S}}(H)$. Once this is done, one can inject the results into estimate (4.11) given by the Fundamental Lemma 4.6; and finally estimate the RHS of (4.11) as a function of the time-step h, hence obtaining the convergence rate.

At the end of Section 5, we discuss the fully second-order discretization scheme when f is allowed to depend only on y and we discuss as well a variance reduction trick for the computation of the involved conditional expectations.

4.3. *Fundamental Lemma for convergence.* The goal of this section is to present a very general but clear result estimating the global error (4.4) of a scheme for BSDE (1.2). Although this type of analysis has already been used in the context of Lipschitz BSDEs [see, e.g., Crisan and Manolarakis (2012), Chassagneux (2012, 2013)], we generalize it to the non-Lipschitz framework we are working with. More precisely, the Fundamental Lemma we present below allows us to cope with schemes which lack stability in the sense of Chassagneux (2013).[10]

4.3.1. *Abstract formulation of a scheme and description of the local error.* In abstract terms, a discretization scheme for a BSDE generates recursively (and backward in time) a family of random variables $\{(Y_i, Z_i)\}_{i=0,\dots,N}$ approximating $\{(Y_{t_i}, \bar{Z}_{t_i})\}_{t_i\in\pi}$ via some operators $\Phi_i : L^2(\mathcal{F}_{i+1}) \times L^2(\mathcal{F}_{i+1}) \to L^2(\mathcal{F}_i) \times L^2(\mathcal{F}_i)$, $i \in \{N-1,\dots,0\}$. One starts with an initial approximation (Y_N, Z_N) and for $i = N-1,\dots,0$ computes $(Y_i, Z_i) := \Phi_i(Y_{i+1}, Z_{i+1})$. [Compare with (4.5)–(4.6) or (4.7)–(4.8).]

Since (Y_i, Z_i) is obtained via Φ_i from the input (Y_{i+1}, Z_{i+1}), we introduce the following notation: for any $i = 0, 1,\dots, N-1$, given a $\mathcal{F}_{t_{i+1}}$-measurable input $(\mathcal{Y}, \mathcal{Z})$, the pair $(Y_{i,(\mathcal{Y},\mathcal{Z})}, Z_{i,(\mathcal{Y},\mathcal{Z})})$ denotes the associated output of $\Phi_i(\mathcal{Y}, \mathcal{Z})$. Writing (Y_i, Z_i) without specifying the input denotes the canonical output of $\Phi_i(Y_{i+1}, Z_{i+1})$, that is, we refer to the family of RV's $\{(Y_i, Z_i)\}_{i=0,\dots,N}$. We introduce as well the notation $\widehat{Y}_i = Y_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})}$ and $\widehat{Z}_i = Z_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})}$ as the output of $\Phi_i(Y_{t_{i+1}}, \bar{Z}_{t_{i+1}})$.

We decompose the local error into two parts: the one-step time-discretization error and the propagation to time t_i of the error from time t_{i+1} (the stability error). So, given $i \in \{0,\dots,N-1\}$, we write

$$\begin{aligned} Y_{t_i} - Y_i &= (Y_{t_i} - \widehat{Y}_i) + (\widehat{Y}_i - Y_i) \\ &= \underbrace{(Y_{t_i} - Y_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})})}_{\text{one-step discretization error}} + \underbrace{(Y_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})} - Y_{i,(Y_{i+1},Z_{i+1})})}_{\text{stability of the scheme}}, \end{aligned}$$

and similarly for Z

$$\begin{aligned} \bar{Z}_{t_i} - Z_i &= (\bar{Z}_{t_i} - \widehat{Z}_i) + (\widehat{Z}_i - Z_i) \\ &= \underbrace{(\bar{Z}_{t_i} - Z_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})})}_{\text{one-step discretization error}} + \underbrace{(Z_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})} - Z_{i,(Y_{i+1},Z_{i+1})})}_{\text{stability of the scheme}}. \end{aligned}$$

We now turn to the question of how to aggregate these errors in order to estimate the global error $\mathrm{ERR}_\pi(Y, Z)$ [see (4.4)].

4.3.2. *The Fundamental Stability Lemma.* The purpose of the Fundamental Lemma below is to formulate in a transparent way the ingredients required to show convergence of $\{(Y_i, Z_i)\}_{i=0,\dots,N}$ to $\{(Y_{t_i}, \bar{Z}_{t_i})\}_{t_i\in\pi}$ in the error criterion (4.4). To start with, we define precisely our concept of stability, generalizing that in Chassagneux (2012) and Chassagneux (2013).

[10] See Definition 2.1 in Chassagneux (2013) with $\zeta_i^Y = \zeta_i^Z = 0$ for $i = 0,\dots, N-1$.

DEFINITION 4.4 (Scheme stability). We say that the numerical scheme $\{(Y_i, Z_i)\}_{i=0,\dots,N}$ is *stable* if for some $\rho > 0$ there exists a constant $c > 0$ such that

$$
\begin{aligned}
&\mathbb{E}\big[|Y_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})} - Y_{i,(Y_{i+1},Z_{i+1})}|^2\big] \\
&\quad + \rho\mathbb{E}\big[|Z_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})} - Z_{i,(Y_{i+1},Z_{i+1})}|^2\big]h \qquad (4.9)\\
&\quad \le (1+ch)\Big(\mathbb{E}\big[|Y_{t_{i+1}} - Y_{i+1}|^2\big] + \frac{\rho}{4}\mathbb{E}\big[|\bar{Z}_{t_{i+1}} - Z_{i+1}|^2\big]h\Big) + \mathbb{E}[H_i],
\end{aligned}
$$

where $H_i \in L^1(\mathcal{F}_i)$, and moreover $\{H_i\}_{i=0,\dots,N-1}$ satisfies

$$
\mathcal{R}^{\mathcal{S}}(H) := \max_{i=0,\dots,N-1} \sum_{j=i}^{N-1} e^{c(j-i)h}\mathbb{E}[H_j] \longrightarrow 0 \qquad \text{as } h \to 0.
$$

The quantity $\mathcal{R}^{\mathcal{S}}(H)$ is called the *stability remainder*.

REMARK 4.5. In the case where f is a globally Lipschitz function, it can be shown for both implicit and explicit schemes that $H_i = 0$ [see Crisan and Manolarakis (2012) or Chassagneux (2013)]. The scheme is then *locally* stable. Our definition of stability allows one to cope with schemes which are not locally stable, as is the case when f is a monotone function with polynomial growth in y, provided we can control the term $\mathcal{R}^{\mathcal{S}}(H)$ (which we do in Section 5). We also point out that it is crucial that in (4.9) we have $\rho > \frac{\rho}{4}$ (compare LHS with RHS). This later allows the use of Gronwall type inequalities (see Lemma A.4).

We now state the Fundamental Lemma which is the basis of the error analysis throughout.

LEMMA 4.6 (Fundamental Lemma). *Assume that the numerical scheme $\{(Y_i, Z_i)\}_{i=0,\dots,N}$ is stable. Denoting the one-step discretization errors for $i = 0, \dots, N-1$ by*

$$
\begin{cases}
\tau_i(Y) := \mathbb{E}\big[|Y_{t_i} - Y_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})}|^2\big] = \mathbb{E}\big[|Y_{t_i} - \widehat{Y}_i|^2\big], \\
\tau_i(Z) := \mathbb{E}\big[|\bar{Z}_{t_i} - Z_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})}|^2 h\big] = \mathbb{E}\big[|\bar{Z}_{t_i} - \widehat{Z}_i|^2 h\big],
\end{cases} \qquad (4.10)
$$

there exists a constant $C = C(\rho, T, c)$ such that

$$
\begin{aligned}
&\big(\mathrm{ERR}_\pi(Y, Z)\big)^2 \\
&\quad \le C\Big\{\mathbb{E}\big[|Y_{t_N} - Y_N|^2\big] + \mathbb{E}\big[|\bar{Z}_{t_N} - Z_N|^2\big]h + \sum_{i=0}^{N-1}\Big(\frac{\tau_i(Y)}{h} + \tau_i(Z)\Big)\Big\} \qquad (4.11)\\
&\qquad + (1+h)\mathcal{R}^{\mathcal{S}}(H).
\end{aligned}
$$

This result states in a rather clear fashion [although $\mathcal{R}^{\mathcal{S}}(H)$ is unknown at this point] what is required in order to have convergence of the numerical scheme. First, one needs a control on the approximation of the terminal conditions [the first two terms in the RHS of (4.11)]. Second, one needs a control on the sum of the one-step time-discretization errors (4.10) [the 3rd term in the RHS of (4.11)]. Third, one need a control on the stability remainder $\mathcal{R}^{\mathcal{S}}(H)$ arising from the scheme stability (4.9) [last term in the RHS of (4.11)]. Of course, the form of $\mathcal{R}^{\mathcal{S}}(H)$ depends on the specific scheme one is handling but in general the error $\mathrm{ERR}_\pi(Y,Z)$ of the scheme is always dominated by (4.11).

The first element will be estimated in Lemma 4.8. The second is the subject of Section 4.6 and the estimate is given in Proposition 4.13. Finally, the study of the stability of the schemes is done in Sections 5 and 6. The convergence rate of the scheme will then follow by estimating further the RHS of (4.11).

PROOF OF LEMMA 4.6. We use throughout the following notation: $\widehat{Y}_i = Y_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})}$, $\widehat{Z}_i = Z_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})}$, $Y_i = Y_{i,(Y_{i+1},Z_{i+1})}$ and $Z_i = Z_{i,(Y_{i+1},Z_{i+1})}$ introduced in Section 4.3.1. We decompose the error as explained above and use Young's inequality to get $|Y_{t_i} - Y_i|^2 \le (1+\frac{1}{h})|Y_{t_i} - \widehat{Y}_i|^2 + (1+h)|\widehat{Y}_i - Y_i|^2$ and $|\bar{Z}_{t_i} - Z_i|^2 h \le 2|\bar{Z}_{t_i} - \widehat{Z}_i|^2 h + 2|\widehat{Z}_i - Z_i|^2 h$.

Using $\rho > 0$ from (4.9) and the definition (4.10) above, it then follows that

$$\begin{aligned}
&\mathbb{E}\big[|Y_{t_i} - Y_i|^2\big] + \frac{\rho}{2}\mathbb{E}\big[|\bar{Z}_{t_i} - Z_i|^2\big]h \\
&\quad \le (1+h)\mathbb{E}\big[|\widehat{Y}_i - Y_i|^2\big] + \rho\mathbb{E}\big[|\widehat{Z}_i - Z_i|^2\big]h + \left(\left(1+\frac{1}{h}\right)\tau_i(Y) + \rho\tau_i(Z)\right).
\end{aligned}$$

Since $\rho \le (1+h)\rho$, by the stability of the scheme [see (4.9)], it follows that

$$\begin{aligned}
&\mathbb{E}\big[|Y_{t_i} - Y_i|^2\big] + \frac{\rho}{2}\mathbb{E}\big[|\bar{Z}_{t_i} - Z_i|^2\big]h \\
&\quad \le (1+h)(1+ch)\left(\mathbb{E}\big[|Y_{t_{i+1}} - Y_{i+1}|^2\big] + \frac{\rho}{4}\mathbb{E}\big[|\bar{Z}_{t_{i+1}} - Z_{i+1}|^2\big]h\right) \\
&\qquad + \left(\left(1+\frac{1}{h}\right)\tau_i(Y) + \rho\tau_i(Z) + (1+h)\mathbb{E}[H_i]\right).
\end{aligned} \tag{4.12}$$

Taking $I_i := |Y_{t_i} - Y_i|^2 + \frac{\rho}{4}|\bar{Z}_{t_i} - Z_i|^2 h$, we have

$$\begin{aligned}
&\mathbb{E}[I_i] + \frac{\rho}{4}\mathbb{E}\big[|\bar{Z}_{t_i} - Z_i|^2\big]h \\
&\quad \le (1+h)(1+ch)\mathbb{E}[I_{i+1}] + \left(\left(1+\frac{1}{h}\right)\tau_i(Y) + \rho\tau_i(Z) + (1+h)\mathbb{E}[H_i]\right),
\end{aligned}$$

and we complete the proof using Lemma A.4. □

4.4. *Discretization of the BSDE.* Let $t_i, t_{i+1} \in \pi$. To approximate the solution (Y, Z) to (1.2), we need two approximations, one for the Y component and one for the Z component. Write (1.2) over the interval $[t_i, t_{i+1}]$ and take $\mathcal{F}_{t_i}$-conditional expectations to obtain [recalling that $\Theta_s = (X_s, Y_s, Z_s)$]

$$Y_{t_i} = \mathbb{E}_{t_i}\left[Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(s, \Theta_s)\,\mathrm{d}s\right]. \tag{4.13}$$

For the Z component, one multiplies (1.2) (written over the interval $[t_i, t_{i+1}]$) by the Brownian increment, $\Delta W_{i+1} := W_{t_{i+1}} - W_{t_i}$, and takes $\mathcal{F}_{t_i}$-conditional expectations to obtain (using Itô's isometry) the implicit formula

$$0 = \mathbb{E}_{t_i}\left[\Delta W_{i+1}\left(Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(s, \Theta_s)\,\mathrm{d}s\right)\right] - \mathbb{E}_{t_i}\left[\int_{t_i}^{t_{i+1}} Z_s\,\mathrm{d}s\right]. \tag{4.14}$$

One now obtains a scheme by approximating the Lebesgue integral via the θ-integration rule (indexed by a parameter $\theta \in [0, 1]$), that is, for some function ψ

$$\int_{t_i}^{t_{i+1}} \psi(s)\,\mathrm{d}s \approx \big[\theta\psi(t_i) + (1-\theta)\psi(t_{i+1})\big](t_{i+1} - t_i), \qquad \theta \in [0, 1].$$

This type of approximation of the integral is generally known to be of first order for $\theta \neq 1/2$ and of higher order for $\theta = 1/2$ (see end of this section). Unfortunately, with the results obtained so far (see Section 3) we are not able to prove the convergence of a general higher order approximation in its full generality; roughly, the issue boils down to obtaining controls on $|\partial^2_{xx} v|$ where v is solution to (2.9). However, under the results of Section 3, we do not even know if $\partial^2_{xx} v$ exists. Under the assumption that f is independent of z, we can prove that the scheme is indeed of higher order (in the y component); the general case is left for future research.

From (4.14) above, we have [compare with (3.23)]

$$\bar{Z}_{t_i} := \frac{1}{h}\mathbb{E}_{t_i}\left[\int_{t_i}^{t_{i+1}} Z_s\,\mathrm{d}s\right] = \frac{1}{h}\mathbb{E}_{t_i}\left[\Delta W_{i+1}\left(Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(s, \Theta_s)\,\mathrm{d}s\right)\right],$$

and we approximate $(Z_s)_{s\in[t_i,t_{i+1}]}$ via $\bar{Z}_{t_i}$ and $\bar{Z}_{t_{i+1}}$ rather than Z_{t_i} or $Z_{t_{i+1}}$. Following the notation for Θ, we denote $\bar{\Theta}_{t_i} := (X_{t_i}, Y_{t_i}, \bar{Z}_{t_i})$ and using the θ-integration rule, it follows

$$\begin{aligned} Y_{t_i} = \mathbb{E}_{t_i}\Big[& Y_{t_{i+1}} + h\big[\theta f(t_i, \bar{\Theta}_{t_i}) + (1-\theta) f(t_{i+1}, \bar{\Theta}_{t_{i+1}})\big] \\ & + \int_{t_i}^{t_{i+1}} R(s)\,\mathrm{d}s\Big], \end{aligned} \tag{4.15}$$

$$\bar{Z}_{t_i} = \mathbb{E}_{t_i}\left[\frac{\Delta W_{i+1}}{h}\left(Y_{t_{i+1}} + (1-\theta) f(t_{i+1}, \bar{\Theta}_{t_{i+1}})h + \int_{t_i}^{t_{i+1}} R(s)\,\mathrm{d}s\right)\right], \tag{4.16}$$

where the error term is, for $s \in [t_i, t_{i+1}]$, defined as $R(s) := \theta R^I(s) + (1-\theta) R^E(s)$ where

$$
\begin{aligned}
R^I(s) &:= f(s, \Theta_s) - f(t_i, \bar{\Theta}_{t_i}) \quad \text{and} \\
R^E(s) &:= f(s, \Theta_s) - f(t_{i+1}, \bar{\Theta}_{t_{i+1}}).
\end{aligned}
\tag{4.17}
$$

REMARK 4.7. For the error analysis here and in the following section, we always understand the set of RVs $\{(Y_{t_i}, \bar{Z}_{t_i})\}_{t_i \in \pi}$ as the true solution of the BSDE on the partition points $t_i \in \pi$ but in the set-up of (4.15) and (4.16). We emphasize that our numerical scheme does not aim at approximating Z itself over π but the family $\{\bar{Z}_{t_i}\}_{t_i \in \pi}$.

The order of the approximation depends on the smoothness of driver f and the properties of the other coefficients. Ignoring the error term R, we find the discretization scheme stated in (4.5)–(4.6). We point out that we aim at first-order schemes, so setting $Z_N = 0$ is not an issue. For a higher order schemes, Z_T needs to be approximated in a more robust fashion, for example, following (3.24), $Z_T = (\nabla_x g)(X_T)\sigma(T, X_T) \approx (\nabla_x g)(X_N)\sigma(T, X_N) = Z_N$ (under the extra assumption that ∇g is Lipschitz).

We can already estimate the error on the terminal conditions, which is the first group of terms in the global error estimate from the Fundamental Lemma 4.6.

LEMMA 4.8. *Let* (HX0), (HY0) *hold. Then there exists a constant c such that* [*recall* (3.23)]

$$
\begin{aligned}
&\mathbb{E}\big[|Y_{t_N} - Y_N|^p\big]^{1/p} \le ch^\gamma \qquad \text{for any } p \ge 2 \quad \text{and} \\
&\mathbb{E}\big[|\bar{Z}_{t_N} - Z_N|^2 h\big] \le ch,
\end{aligned}
\tag{4.18}
$$

where γ is the order of the approximation $\{X_i\}_{i=0,\dots,N}$ of X [*according to* (4.1)].

Assume that $g \in C_b^1$ and that ∇g is Lipschitz continuous. Define $Z_N := (\nabla_x g)(X_N)\sigma(T, X_N)$ then $\mathbb{E}[|\bar{Z}_{t_N} - Z_N|^2 h] \le ch^2$.

PROOF. The error estimate on Y_{t_N} results from the Lipschitz regularity of g and the estimate on $\mathbb{E}[|X_{t_N} - X_N|^2]$ given by (4.1). For the error estimate on Z, we have $Z_N = 0$, and $\bar{Z}_{t_N} = Z_T$, which in turn implies $\mathbb{E}[|\bar{Z}_{t_N} - Z_N|^2 h] = \mathbb{E}[|Z_T|^2]h \le ch$ where we have used (3.24).

In the case where $g \in C_b^1$ and ∇g is Lipschitz, the estimate follows easily using that $\bar{Z}_T = Z_T = \nabla g(X_T)\sigma(T, X_T)$ and using the Lipschitz property of ∇g and σ, the Cauchy–Schwarz inequality and (4.1). □

4.5. *Existence and local estimates for the general θ-scheme.* In this subsection, we start the study of the θ-scheme (4.5)–(4.6) by analyzing *one step* of it, that is, going from time t_{i+1} to t_i. To simplify notation, we define $f_{i+1} := f(t_{i+1}, X_{i+1}, Y_{i+1}, Z_{i+1})$ and $A_{i+1} := Y_{i+1} + (1-\theta) f_{i+1} h$.

Along with (HX0) and (HY0), we make the temporary assumption that $Y_{i+1}, Z_{i+1}, f_{i+1} \in L^2$ (this integrability assumption is clearly satisfied by Y_N, Z_N and f_N) and analyze how, when $\theta > 0$, this integrability carries on to the next time step.

Note that for $\theta = 0$ (i.e., the explicit case) the scheme step is well defined as Y_i and Z_i can be easily computed. For $\theta > 0$, there is no issue in defining Z_i from (4.6), but unlike in the Lipschitz case, it is not immediate that the solution Y_i to the implicit equation (4.5) exists. We need to show first that there exists a unique Y_i solving $Y_i = \mathbb{E}_i[A_{i+1}] + \theta f(t_i, X_i, Y_i, Z_i)h$, where $\mathbb{E}_i[A_{i+1}]$, X_i and Z_i are already known. This follows from Theorem 26.A in Zeidler [(1990), page 557]. Define (almost surely) the map $F : y \mapsto y - \theta f(t_i, X_i(\omega), y, Z_i(\omega))h$. This map is strongly monotone (increasing) in the sense of Definition 25.2 in Zeidler (1990), that is, there exists a $\mu > 0$ such that for all y, y',

$$\langle y' - y, F(y') - F(y)\rangle \geq \mu |y' - y|^2.$$

Indeed, from (HY0) and Remark 2.1 we have

$$\langle y' - y, F(y') - F(y)\rangle \geq (1 - \theta L_y h)|y' - y|^2,$$

so if $h < 1/(\theta L_y)$ we can take $\mu = (1 - \theta L_y h) > 0$. This (almost surely) guarantees the existence of a unique $Y_i(\omega) = F^{-1}[\mathbb{E}_i[(A_{i+1})](\omega)]$, as needed. By the monotonicity of F, Y_i can be quickly computed using, for example, Newton–Raphson-type methods. Now, Y_i so defined is only an $\mathcal{F}_i$-measurable random variable.[11]

The following proposition guarantees that if $\theta > 0$, the pair (Y_i, Z_i) and the term f_i are square integrable provided the corresponding random variables at t_{i+1} also are. So for every N, by iteration, (Y_i, Z_i) is well defined for $i = N-1, \dots, 0$. For $\theta \geq 1/2$, this estimate also leads to a uniform bound, as will become clear in the next section (Proposition 5.1).

PROPOSITION 4.9. *Let* (HX0), (HY0) *hold*, $\theta \in [0, 1]$ *and take* $h \leq \min\{1, [4\theta(L_y + 3d\theta L_z^2)]^{-1}\}$. *Then there exists a constant* c *such that for any*

[11]The previous explanation only justified the existence of Y_i as a function from Ω to $\mathbb{R}^k$. To obtain that it is measurable, one should rather consider the map $G : (a, y) \mapsto (a, y - \theta f(t_i, a, y)h)$, where $a = (x, z) \in \mathbb{R}^{d \times k \times d}$ and $f(t, a, y) = f(t, x, y, z)$. It is again seen to be strongly monotonous, so it is invertible and Theorem 26.A in Zeidler (1990) asserts that G^{-1} is continuous (Lipschitz in fact), hence measurable.

$i \in \{0, \dots, N-1\}$

$$
\begin{aligned}
&|Y_i|^2 + \frac{1}{2d}|Z_i|^2 h + 2\theta^2 |f_i|^2 h^2 \\
&\quad \le (1+ch)\mathbb{E}_i\left[|Y_{i+1}|^2 + \frac{1}{8d}|Z_{i+1}|^2 h\right] + ch \\
&\qquad + c\big(|X_i|^2 + \mathbb{E}_i\big[|X_{i+1}|^2\big]\big)h + 2(1-\theta)^2 \mathbb{E}_i\big[|f_{i+1}|^2\big]h^2.
\end{aligned} \tag{4.19}
$$

PROOF OF PROPOSITION 4.9. Let $i \in \{0, \dots, N-1\}$. First, we estimate Z_i. The martingale property of ΔW_{i+1} yields

$$
Z_i h = \mathbb{E}_i[\Delta W_{i+1} A_{i+1}] = \mathbb{E}_i\big[\Delta W_{i+1}\big(A_{i+1} - \mathbb{E}_i[A_{i+1}]\big)\big]. \tag{4.20}
$$

By the Cauchy–Schwarz inequality,

$$
|Z_i|^2 h \le d\big\{\mathbb{E}_i\big[A_{i+1}^2\big] - \mathbb{E}_i[A_{i+1}]^2\big\}. \tag{4.21}
$$

We now proceed with the estimation of Y_i. We first rewrite

$$
Y_i = \mathbb{E}_i[A_{i+1}] + \theta f_i h \quad \Longleftrightarrow \quad Y_i - \theta f_i h = \mathbb{E}_i[A_{i+1}]
$$

and then square both sides of the RHS of the above equivalence to obtain

$$
|Y_i|^2 = \mathbb{E}_i[A_{i+1}]^2 + 2\theta\langle Y_i, f_i\rangle h - \theta^2 |f_i|^2 h^2.
$$

This simple manipulation allows us to take advantage of the monotonicity of f [see (2.1)] and will be reused frequently. By the estimate of Remark 2.1, with an $\alpha > 0$ to be chosen later, the previous equality leads to

$$
|Y_i|^2 \le \mathbb{E}_i[A_{i+1}]^2 + 2\theta(L_y + \alpha)|Y_i|^2 h + \theta B(i,\alpha) + \frac{3\theta L_z^2}{2\alpha}|Z_i|^2 h - \theta^2 |f_i|^2 h^2,
$$

where $B(i,\alpha) := (3L^2 h + 3L_x^2 |X_i|^2 h)/(2\alpha)$. Now, for $\epsilon = 1/d$, we combine the above estimate with (4.21) to obtain

$$
\begin{aligned}
|Y_i|^2 + \epsilon |Z_i|^2 h &\le (1 - \epsilon d)\mathbb{E}_i[A_{i+1}]^2 + \epsilon d \mathbb{E}_i\big[A_{i+1}^2\big] \\
&\quad + 2\theta(L_y + \alpha)|Y_i|^2 h + \frac{3\theta L_z^2}{2\alpha}|Z_i|^2 h + \theta B(i,\alpha) - \theta^2 |f_i|^2 h^2.
\end{aligned}
$$

Reorganizing the terms leads to

$$
\begin{aligned}
&\big(1 - 2\theta(L_y + \alpha)h\big)|Y_i|^2 + \left(\epsilon - \frac{3\theta L_z^2}{2\alpha}\right)|Z_i|^2 h \\
&\quad \le \mathbb{E}_i\big[A_{i+1}^2\big] + \theta B(i,\alpha) - \theta^2 |f_i|^2 h^2.
\end{aligned} \tag{4.22}
$$

Using again Remark 2.1 with $\alpha' > 0$, we obtain

$$
\begin{aligned}
A_{i+1}^2 &\le |Y_{i+1}|^2 + (1-\theta)2\big(L_y + \alpha'\big)|Y_{i+1}|^2 h \\
&\quad + (1-\theta)\frac{3L_z^2}{2\alpha'}|Z_{i+1}|^2 h + (1-\theta)B\big(i+1,\alpha'\big) + (1-\theta)^2 |f_{i+1}|^2 h^2,
\end{aligned}
$$

which in turns leads to

$$(4.23)\qquad \begin{aligned} &(1-2\theta(L_y+\alpha)h)|Y_i|^2+\left(\epsilon-\frac{3\theta L_z^2}{2\alpha}\right)|Z_i|^2h \\ &\quad\le \big(1+(1-\theta)2\big(L_y+\alpha'\big)h\big)\mathbb{E}_i\big[|Y_{i+1}|^2\big] \\ &\qquad+(1-\theta)\frac{3L_z^2}{2\alpha'}\mathbb{E}_i\big[|Z_{i+1}|^2\big]h+H_i^\theta \\ &\qquad+\theta B(i,\alpha)+(1-\theta)\mathbb{E}_i\big[B\big(i+1,\alpha'\big)\big], \end{aligned}$$

where

$$(4.24)\qquad H_i^\theta:=(1-\theta)^2\mathbb{E}_i\big[|f_{i+1}|^2\big]h^2-\theta^2|f_i|^2h^2.$$

Now, we choose $\alpha=3d\theta L_z^2$ (so that $\epsilon-\frac{3\theta L_z^2}{2\alpha}=\frac{1}{2d}$) and $\alpha'=24d(1-\theta)L_z^2$ [so that $(1-\theta)\frac{3L_z^2}{2\alpha'}\le\frac{1}{16d}$]. Since $h\le\min\{1,[4\theta(L_y+3d\theta L_z^2)]^{-1}\}$ it is true that $2\theta(L_y+\alpha)h\le 1/2$. We also observe that for $x\in[0,1/2]$, $1\le 1/(1-x)\le 1+2x\le 2$ and as a consequence

$$\begin{aligned} &|Y_i|^2+\frac{1}{2d}|Z_i|^2h \\ &\quad\le\big(1+4\theta(L_y+\alpha)h\big)\big(1+2(1-\theta)\big(L_y+\alpha'\big)h\big)\mathbb{E}_i\big[|Y_{i+1}|^2\big] \\ &\qquad+\frac{1}{8d}\mathbb{E}_i\big[|Z_{i+1}|^2\big]h+2\theta B(i,\alpha)+2(1-\theta)\mathbb{E}_i\big[B\big(i+1,\alpha'\big)\big]+2H_i^\theta. \end{aligned}$$

Defining $c:=4\theta(L_y+\alpha)+2(1-\theta)(L_y+\alpha')+8\theta(L_y+\alpha)(1-\theta)(L_y+\alpha')$, we clearly have

$$\big(1+4\theta(L_y+\alpha)h\big)\big(1+2(1-\theta)\big(L_y+\alpha'\big)h\big)\le 1+ch.$$

We can now conclude to the announced estimate

$$(4.25)\qquad \begin{aligned} &|Y_i|^2+\frac{1}{2d}|Z_i|^2h \\ &\quad\le(1+ch)\left(\mathbb{E}_i\big[|Y_{i+1}|^2\big]+\frac{1}{8d}\mathbb{E}_i\big[|Z_{i+1}|^2\big]h\right) \\ &\qquad+2\theta B(i,\alpha)+2(1-\theta)\mathbb{E}_i\big[B\big(i+1,\alpha'\big)\big]+2H_i^\theta, \end{aligned}$$

provided one passes the term $-2\theta^2|f_i|^2h^2$ in $2H_i^\theta$ to the LHS. This completes the proof. □

4.6. *Local time-discretization error.* As announced in Sections 4.2 and 4.3, we now proceed to estimating the one-step discretization errors $\tau_i(Y)$ and $\tau_i(Z)$ [see (4.10) for the definition], and then their sum. We thus obtain an estimate for the second group of terms in estimate (4.11), which is summarized in Proposition 4.13.

We follow the notation of Section 4.3 and write, for $i = 0, 1, \dots, N-1$, $\widehat{Y}_i = Y_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})}$ and $\widehat{Z}_i = Z_{i,(Y_{t_{i+1}},\bar{Z}_{t_{i+1}})}$; that is, $(\widehat{Y}_i, \widehat{Z}_i)$ is the solution to

$$\begin{aligned}\widehat{Y}_i &= \mathbb{E}_{t_i}[Y_{t_{i+1}} + (1-\theta) f(t_{i+1}, X_{i+1}, Y_{t_{i+1}}, \bar{Z}_{t_{i+1}})h] \\ &\quad + \theta f(t_i, X_i, \widehat{Y}_i, \widehat{Z}_i)h,\end{aligned} \tag{4.26}$$

$$\widehat{Z}_i = \mathbb{E}_{t_i}\left[\frac{\Delta W_{i+1}}{h}(Y_{t_{i+1}} + (1-\theta) f(t_{i+1}, X_{i+1}, Y_{t_{i+1}}, \bar{Z}_{t_{i+1}})h)\right]. \tag{4.27}$$

REMARK 4.10. We know from Proposition 4.9 that, under the assumption $h \le \min\{1, [4\theta(L_y + 3d\theta L_z^2)]^{-1}\}$, the RV's $\{(\widehat{Y}_i, \widehat{Z}_i)\}_{i=0,\dots,N}$ are well defined and square integrable. Furthermore, estimate (4.19), together with the growth assumption on f in (HY0), (4.1) for X_{i+1}, Theorem 2.2 for $Y_{t_{i+1}}$ and Corollary 3.6 for $\bar{Z}_{t_{i+1}}$, guarantee immediately that for any $p \ge 2$, there exists a constant c such that

$$\sup_{N\in\mathbb{N}} \max_{i=0,\dots,N} \mathbb{E}[|\widehat{Y}_i|^p] \le c. \tag{4.28}$$

This fact will be needed later in Section 5 (in Lemma 5.3).

The next result estimates the one-step discretization errors $\tau_i(Y)$ and $\tau_i(Z)$ of the approximation in terms of the error process R [as defined in (4.17)]. Afterward, we discuss the behavior of R itself.

LEMMA 4.11. *Let* (HX0) *and* (HY0) *hold and assume that* $h \le 1/(4\theta L_y)$. *Then for any* $\theta \in [0, 1]$ *there exists a constant* c *such that for any* $i \in \{0, \dots, N-1\}$

$$\mathbb{E}[|Y_{t_i} - \widehat{Y}_i|^2 + |\bar{Z}_{t_i} - \widehat{Z}_i|^2 h] \le c\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} R(s)\,\mathrm{d}s\right)^2\right] + cL_x^2 \mathrm{ERR}_\pi(X)^2 h^2.$$

PROOF. Let $i \in \{0, \dots, N-1\}$. Recalling (4.16), (4.27) and the definition $\bar{\Theta}_{t_i} := (X_{t_i}, Y_{t_i}, \bar{Z}_{t_i})$ we have

$$\begin{aligned}\bar{Z}_{t_i} - \widehat{Z}_i = \mathbb{E}_i\biggl[\frac{\Delta W_{i+1}}{h}\biggl(&(1-\theta)[f(t_{i+1}, \bar{\Theta}_{t_{i+1}}) - f(t_{i+1}, X_{i+1}, Y_{t_{i+1}}, \bar{Z}_{t_{i+1}})]h \\ &+ \int_{t_i}^{t_{i+1}} R(s)\,\mathrm{d}s\biggr)\biggr],\end{aligned}$$

which by the Cauchy–Schwarz inequality and the Lipschitz property of the map $x \mapsto f(\cdot, x, \cdot, \cdot)$ leads to

$$h|\bar{Z}_{t_i} - \widehat{Z}_i|^2 \le 2d\mathbb{E}_i\left[\left(\int_{t_i}^{t_{i+1}} R_u\,\mathrm{d}u\right)^2\right] + 2d(1-\theta)^2 L_x^2 \mathbb{E}_i[|X_{t_{i+1}} - X_{i+1}|^2]h^2.$$

For the Y-part, similarly by recalling (4.15) and (4.26), we have

$$
\begin{aligned}
&Y_{t_i} - \widehat{Y}_i \\
&\quad = \mathbb{E}_i\bigg[\int_{ti}^{t_{i+1}} R(s)\,\mathrm{d}s + (1-\theta)\big(f(t_{i+1}, \bar{\Theta}_{t_{i+1}}) - f(t_{i+1}, X_{i+1}, Y_{t_{i+1}}, \bar{Z}_{t_{i+1}})\big)h\bigg] \\
&\qquad + \theta\big(f(t_i, \bar{\Theta}_{t_i}) - f(t_i, X_i, \widehat{Y}_i, \widehat{Z}_i)\big)h \\
&\quad = \mathbb{E}_i\bigg[\int_{t_i}^{t_{i+1}} R(s)\,\mathrm{d}s + (1-\theta)\big(f(t_{i+1}, \bar{\Theta}_{t_{i+1}}) - f(t_{i+1}, X_{i+1}, Y_{t_{i+1}}, \bar{Z}_{t_{i+1}})\big)h\bigg] \\
&\qquad + \theta\big(f(t_i, X_{t_i}, Y_{t_i}, \bar{Z}_{t_i}) - f(t_i, X_i, Y_{t_i}, \widehat{Z}_i)\big)h \\
&\qquad + \theta\big(f(t_i, X_i, Y_{t_i}, \widehat{Z}_i) - f(t_i, X_i, \widehat{Y}_i, \widehat{Z}_i)\big)h.
\end{aligned}
$$

To obtain the estimate for $|Y_{t_i} - \widehat{Y}_i|^2$, similarly as in the proof of Proposition 4.9, we pass the last term in the RHS to the LHS, square both sides, expand the square on the LHS, pass the cross term to the RHS and dominate it on the RHS using (2.1). By collecting only the convenient terms in the LHS and using assumption (HY0) on the RHS, we get

$$
\begin{aligned}
|Y_{t_i} - \widehat{Y}_i|^2 &\le 3\mathbb{E}_i\bigg[\int_{t_i}^{t_{i+1}} R(s)\,\mathrm{d}s\bigg]^2 + 6\theta^2 L_z^2 |\bar{Z}_{t_i} - \widehat{Z}_i|^2 h^2 + 2\theta L_y |Y_{t_i} - \widehat{Y}_i|^2 h \\
&\qquad + 6\theta^2 L_x^2 |X_{t_i} - X_i|^2 h^2 + 3(1-\theta)^2 L_x^2 \mathbb{E}_i\big[|X_{t_{i+1}} - X_{i+1}|^2\big] h^2,
\end{aligned}
$$

which implies, using the estimate for $|\bar{Z}_{t_i} - \widehat{Z}_i|^2$, that

$$
\begin{aligned}
&(1 - 2\theta L_y h)|Y_{t_i} - \widehat{Y}_i|^2 \\
&\qquad \le (3 + 12d\theta^2 L_z^2 h)\mathbb{E}_i\bigg[\bigg(\int_{t_i}^{t_{i+1}} R(s)\,\mathrm{d}s\bigg)^2\bigg] + 6\theta^2 L_x^2 |X_{t_i} - X_i|^2 h^2 \\
&\qquad\quad + 3(1-\theta)^2 L_x^2 (1 + 4d\theta^2 L_z^2 h)\mathbb{E}_i\big[|X_{t_{i+1}} - X_{i+1}|^2\big] h^2.
\end{aligned}
$$

Noting that h is such that $2\theta L_y h \le 1/2$ and by combining the estimates for $|Y_{t_i} - \widehat{Y}_i|^2$ and $|\bar{Z}_{t_i} - \widehat{Z}_i|^2$ the sought result follows after taking expectations and using (4.1) for X. □

We now estimate the integral of the error function R [see (4.17)].

LEMMA 4.12. *Let* (HX0), (HY0$_{\mathrm{loc}}$) *hold. Then there exists* $c > 0$ *such that, for any* $\theta \in [0, 1]$ *and* $i \in \{0, \dots, N-1\}$,

$$
\begin{aligned}
&\mathbb{E}\bigg[\bigg(\int_{t_i}^{t_{i+1}} R(s)\,\mathrm{d}s\bigg)^2\bigg] \\
&\qquad \le c L_t^2 h^3 + c L_x^2 \mathrm{REG}_\pi(X)^2 h^2 + c L_y \mathrm{REG}_\pi(Y)^2 h^2 \\
&\qquad\quad + c L_z^2 \mathbb{E}\bigg[\int_{t_i}^{t_{i+1}} |Z_s - \bar{Z}_{t_i}|^2\,\mathrm{d}s + \int_{t_i}^{t_{i+1}} |Z_s - \bar{Z}_{t_{i+1}}|^2\,\mathrm{d}s\bigg] h.
\end{aligned}
$$

PROOF. Following from (4.17), we estimate R via R^I and R^E: using ($\mathrm{HY0}_{\mathrm{loc}}$), Cauchy–Schwarz's inequality and Fubini's theorems we have [recall that $\Theta = (X, Y, Z)$ and $\bar{\Theta}_{t_i} = (X_{t_i}, Y_{t_i}, \bar{Z}_{t_i})$]

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\int_{t_i}^{t_{i+1}} R^I(s)\,\mathrm{d}s\Big)^2\Big] \\
&\quad= \mathbb{E}\Big[\Big(\int_{t_i}^{t_{i+1}} \big[f(s,\Theta_s) \pm f(s, X_s, Y_{t_i}, Z_s) - f(t_i, \bar{\Theta}_{t_i})\big]\,\mathrm{d}s\Big)^2\Big] \\
&\quad\le 2h\mathbb{E}\Big[\int_{t_i}^{t_{i+1}} 3L_y^2\big(1 + |Y_s|^{2(m-1)} + |Y_{t_i}|^{2(m-1)}\big)|Y_s - Y_{t_i}|^2\,\mathrm{d}s + \alpha_i\Big] \\
&\quad\le 2h\Big(\int_{t_i}^{t_{i+1}} L_y^2\mathbb{E}\big[3\big(1 + |Y_s|^{4(m-1)} + |Y_{t_i}|^{4(m-1)}\big)\big]^{1/2}\mathbb{E}\big[|Y_s - Y_{t_i}|^4\big]^{1/2}\,\mathrm{d}s \\
&\qquad\qquad + \mathbb{E}[\alpha_i]\Big),
\end{aligned}$$

where $\alpha_i = 3\int_{t_i}^{t_{i+1}}[L_t^2|s - t_i| + L_x^2|X_s - X_{t_i}|^2 + L_z^2|Z_s - \bar{Z}_{t_i}|^2]\,\mathrm{d}s$.

Using Theorem 2.2 to deal with the Y component, this yields the estimate

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\int_{t_i}^{t_{i+1}} R^I(s)\,\mathrm{d}s\Big)^2\Big] \\
&\qquad\le 3L_t^2h^3 + 6L_x^2\mathrm{REG}_\pi(X)^2h^2 + 18cL_y^2\mathrm{REG}_\pi(Y)^2h^2 \\
&\qquad\quad + 6L_z^2\mathbb{E}\Big[\int_{t_i}^{t_{i+1}} |Z_s - \bar{Z}_{t_i}|^2\,\mathrm{d}s\Big]h.
\end{aligned}$$

Similar arguments allow a similar estimate for R^E but with terms t_{i+1}, $X_{t_{i+1}}$, $Y_{t_{i+1}}$ and $\bar{Z}_{t_{i+1}}$ instead of t_i, X_{t_i}, Y_{t_i} and $\bar{Z}_{t_i}$. $\square$

The trapezoidal integration case. Here, we refine the analysis of the local discretization error from Lemma 4.12 for the case $\theta = 1/2$ in order to obtain better global error estimates. We drop the Z-dependence in f due to lacking regularity results. Approximation (4.6) is found by approximating the last integral on the RHS of (4.14) by a first-order approximation and so it should be clear that at best the overall order of the scheme would be one (in the next section we propose a candidate for higher order approximation of Z). We point out nonetheless that many reaction–diffusion equations have a driver f that only depends on Y. For ease of the presentation, we also assume that f does not depend on the forward process X and omit the time dependence (these can be easily extended).

We write, similarly to (4.15),

$$\int_{t_i}^{t_{i+1}} f(Y_s)\,\mathrm{d}s = \frac{h}{2}\big[f(Y_{t_i}) + f(Y_{t_{i+1}})\big] + \int_{t_i}^{t_{i+1}} R(s)\,\mathrm{d}s,$$

with

$$R(s) := f(Y_s) - \tfrac{1}{2}\big[f(Y_{t_i}) + f(Y_{t_{i+1}})\big],$$

where, using integration by parts, it can be shown [see Süli and Mayers (2003)] that

$$\mathbb{E}\bigg[\bigg(\int_{t_i}^{t_{i+1}} R(s)\,\mathrm{d}s\bigg)^2\bigg] \le \frac{h^6}{12^2}\mathbb{E}\Big[\sup_{t_i \le t \le t_{i+1}} \big|\partial^2_{yy} f(Y_t)\big|^2\Big]. \tag{4.29}$$

Hence, in the special case where the driver of FBSDE under consideration does not depend on the process $(Z_t)_{0\le t\le T}$ we can take full advantage of trapezoidal integration rule provided that the second derivatives of f in the y variable has polynomial growth, so that there exists a constant c for which

$$\max_{t_i, t_{i+1}\in\pi} \mathbb{E}\Big[\sup_{t_i \le t \le t_{i+1}} \big|\partial^2_{yy} f(Y_t)\big|^2\Big] \le c.$$

The result on the sum of local errors. In view of the above lemmas [as well as estimate (4.1) and the path-regularity Theorem 3.5], we can state the following estimates on the sum of the one-step discretization errors, as appearing in the global error estimate (4.11) of Lemma 4.6.

PROPOSITION 4.13. *Let* (HX0), (HY0$_{\text{loc}}$) *hold and* $h \le \min\{1, [4\theta(L_y + 3d\theta L_z^2)]^{-1}\}$. *For the scheme* (4.5)–(4.6) *we have the following local error estimates*:

(i) *For any* $\theta \in [0,1]$ $\exists c > 0$ *such that* $\sum_{i=0}^{N-1} \frac{\tau_i(Y)}{h} \le ch$ *and* $\sum_{i=0}^{N-1} \tau_i(Z) \le ch^2$.

(ii) *Take* $\theta = 1/2$ *and scheme* (4.5). *Assume additionally that* $f \in C^2$ *does not depend on* (t, x, z) *and* $\partial^2_{yy} f$ *has at most polynomial growth*, *then there exists* $c > 0$ *such that* $\sum_{i=0}^{N-1} \frac{\tau_i(Y)}{h} \le ch^4$.

PROOF. Recall the definition of $\tau_i(Y)$ and $\tau_i(Z)$ given in (4.10). The proof of case (i) is simple: inject in the estimate of Lemma 4.11 that of Lemma 4.12 and then sum over $i = 0$ to $i = N-1$. On the resulting inequality,

$$\sum_{i=0}^{N-1} \tau_i(Y) + \tau_i(Z) \le cL_t^2 h^2 + cL_x^2 \mathrm{REG}_\pi(X)^2 h + cL_y^2 \mathrm{REG}_\pi(Y)^2 h + cL_z^2 \mathrm{REG}_\pi(Z)^2 h + cL_x^2 \mathrm{ERR}_\pi(X)^2 h,$$

apply (4.1) for $\mathrm{ERR}_\pi(X)$, the path-regularity result (4.3) for $\mathrm{REG}_\pi(X)$, and the path-regularity Theorem 3.5 for $\mathrm{REG}_\pi(Y)$ and $\mathrm{REG}_\pi(Z)$. Under (HX0) and (HY0$_{\text{loc}}$) the resulting inequality is $\sum_{i=0}^{N-1}(\tau_i(Y) + \tau_i(Z)) \le ch^2$. The statement's inequalities now follows.

For the proof of case (ii), remark that (4.26) is now independent of Z, and hence using Lemma 4.11 in combination with (4.29) instead of Lemma 4.12 yields the result. □

REMARK 4.14. Under the assumption that f only depends on y (i.e., take $L_t = L_x = L_z = 0$) the methodology used above yields that the first terms in the global error $\mathrm{ERR}_\pi(Y, Z)$ [see (4.11)] is controlled only by $\mathrm{ERR}_\pi(X)$ and $\mathrm{REG}_\pi(Y)$. The term $\mathrm{REG}_\pi(Y)$ follows from the sum of the local discretization errors, as can be seen from above, while $\mathrm{ERR}_\pi(X)$ follows from the approximation of the terminal condition.

These abstract estimates suggest that under stronger regularity assumptions on f [stronger than $(\mathrm{HY0}_{\mathrm{loc}})$], one may improve the estimates on $\tau(Y)$ and therefore obtain a higher convergence rate. Such developments are left for future research.

5. Convergence of the implicit-leaning schemes ($1/2 \le \theta \le 1$). In this section, we complete the convergence proof of the theta scheme (4.5)–(4.6) for $\theta \in [1/2, 1]$ as stated in Theorem 4.2. In view of the Fundamental Lemmas 4.6, 4.8 and Proposition 4.13, what remains to study is the stability of the scheme and estimate $\mathcal{R}^{\mathcal{S}}(H)$.

5.1. *Integrability for the θ-scheme, for $1/2 \le \theta \le 1$.* We now show that for $\theta \ge 1/2$ the scheme cannot explode as h vanishes. These L^p estimates will be useful in obtaining the stability of the scheme.

PROPOSITION 5.1. *Let* (HX0), (HY0) *hold, and* $h \le \min\{1, [4\theta(L_y + 3d\theta L_z^2)]^{-1}\}$ *and let* $\theta \in [1/2, 1]$. *Then for any* $p \ge 1$, *there exists a constant* c *such that*

$$\max_{i=0,\dots,N} \mathbb{E}\big[|Y_i|^{2p}\big] + \sum_{i=0}^{N-1} \mathbb{E}\big[\big(|Z_i|^2 h\big)^p\big] \le c\big(1 + \mathbb{E}\big[|X_N|^{2mp}\big]\big).$$

PROOF. Take $i \in \{0, \dots, N-1\}$ and define the quantity $I_i := |Y_i|^2 + \frac{1}{8d}|Z_i|^2 h + \theta^2 |f(t_i, X_i, Y_i, Z_i)|^2 h^2$. By Proposition 4.9 and that $(1-\theta)^2 \le \theta^2$, for $\theta \in [1/2, 1]$, we have for $\beta_i := c + c(|X_i|^2 + |X_{i+1}|^2)$ the inequality

$$I_i + \frac{3}{8d}|Z_i|^2 h \le e^{ch}\mathbb{E}_i[I_{i+1}] + \mathbb{E}_i[\beta_i]h. \tag{5.1}$$

As a consequence of Lemma A.4, we know that, since $\beta_j \ge 0$,

$$I_i + \frac{3}{8d}\mathbb{E}_i\Bigg[\sum_{j=i}^{N-1} |Z_j|^2 h\Bigg] \le e^{cT}\Bigg(\mathbb{E}_i[I_N] + \sum_{j=i}^{N-1} \mathbb{E}_i[\beta_j]h\Bigg),$$

in particular, using Jensen's inequality, we obtain further

$$|I_i|^p \le 2^{p-1} e^{cpT} \left(\mathbb{E}_i\big[|I_N|^p\big] + (Nh)^{p-1} \sum_{j=0}^{N-1} \mathbb{E}_i\big[|\beta_j|^p\big] h \right).$$

This then implies, thanks to (HY0),

$$\max_{i=0,\dots,N} \mathbb{E}\big[|I_i|^p\big] \le c\big(1 + \mathbb{E}\big[|X_N|^{2mp}\big]\big)$$
$$\Rightarrow \quad \max_{i=0,\dots,N} \mathbb{E}\big[|Y_i|^{2p}\big] \le c\big(1 + \mathbb{E}\big[|X_N|^{2mp}\big]\big).$$

From (5.1), we also have

$$\begin{aligned} I_i^p &+ \left(\frac{3}{8d}\right)^p \big(|Z_i|^2 h\big)^p \\ &\le \left(I_i + \frac{3}{8d}|Z_i|^2 h\right)^p \\ &\le e^{cph} \mathbb{E}_i\big[I_{i+1}^p\big] + \sum_{j=1}^{p} \binom{p}{j} \big(e^{ch} \mathbb{E}_i[I_{i+1}]\big)^{p-j} \big(\mathbb{E}_i[\beta_i] h\big)^j, \end{aligned}$$

so that, applying again Lemma A.4 along with Hölder's and Jensen's inequalities we have

$$\begin{aligned} &\left(\frac{3}{8d}\right)^p \mathbb{E}\left[\sum_{i=0}^{N-1} \big(|Z_i|^2 h\big)^p\right] \\ &\quad\le e^{cpT} \mathbb{E}\big[|I_N|^p\big] + \sum_{i=0}^{N-1} e^{cih} \sum_{j=1}^{p} \binom{p}{j} \mathbb{E}\big[\big(e^{ch} \mathbb{E}_i[I_{i+1}]\big)^{p-j} \big(\mathbb{E}_i[\beta_i] h\big)^j\big] \\ &\quad\le e^{cpT} \mathbb{E}\big[|I_N|^p\big] + e^{cpT} \sum_{i=0}^{N-1} \sum_{j=1}^{p} \binom{p}{j} \big(\mathbb{E}\big[|I_{i+1}|^p\big]\big)^{(p-j)/p} \big(\mathbb{E}\big[|\beta_i|^p\big]\big)^{j/p} h \\ &\quad\le e^{cpT} \mathbb{E}\big[|I_N|^p\big] \\ &\qquad + e^{cpT} T \sum_{j=1}^{p} \binom{p}{j} \Big(\max_{i=0,\dots,N} \mathbb{E}\big[|I_{i+1}|^p\big]\Big)^{(p-j)/p} \Big(\max_{i=0,\dots,N} \mathbb{E}\big[|\beta_i|^p\big]\Big)^{j/p}. \end{aligned}$$

Due to (HY0) and the previous estimates we arrive, as required, at

$$\mathbb{E}\left[\sum_{i=0}^{N-1} \big(|Z_i|^2 h\big)^p\right] \le c\big(1 + |X_N|^{2mp}\big).$$

□

5.2. *Stability of the θ-scheme for* $1/2 \le \theta \le 1$. We now study the stability of the scheme in the sense of (4.9). We fix $i \in \{0, \dots, N-1\}$ and estimate the distance between the outputs $(\widehat{Y}_i, \widehat{Z}_i)$ [see (4.26)–(4.27)] and (Y_i, Z_i) [see (4.5)–(4.6)] as a function of the distance between the inputs $(Y_{t_{i+1}}, \bar{Z}_{t_{i+1}})$ and (Y_{i+1}, Z_{i+1}).

We use the notation $\delta Y_{i+1} = Y_{t_{i+1}} - Y_{i+1}$, $\delta Z_{i+1} := \bar{Z}_{t_{i+1}} - Z_{i+1}$, as well as

$$\delta f_{i+1} = f(t_{i+1}, X_{i+1}, Y_{t_{i+1}}, \bar{Z}_{t_{i+1}}) - f(t_{i+1}, X_{i+1}, Y_{i+1}, Z_{i+1})$$

and

$$\delta A_{i+1} = \delta Y_{i+1} + (1-\theta)\delta f_{i+1} h.$$

Then, denoting by $\widehat{\delta Y_i} = \widehat{Y}_i - Y_i$, $\widehat{\delta Z_i} = \widehat{Z}_i - Z_i$ and $\delta \widehat{f}_i = f(t_i, X_i, \widehat{Y}_i, \widehat{Z}_i) - f(t_i, X_i, Y_i, Z_i)$, we can write that [compare with (4.26), (4.27), (4.5) and (4.6)]

$$\widehat{\delta Y_i} = \mathbb{E}_i[\delta A_{i+1}] + \theta \delta \widehat{f}_i h \quad \text{and} \quad \widehat{\delta Z_i} = \mathbb{E}_i\left[\frac{1}{h}\Delta W_{i+1} \delta A_{i+1}\right].$$

PROPOSITION 5.2. *Let* (HX0) *and* (HY0) *hold. Then there exists a constant* c *for any* $i \in \{0, \dots, N-1\}$ *and* $h \le \min\{1, [4\theta(L_y + d\theta L_z^2)]^{-1}\}$ *such that*

$$|\widehat{\delta Y_i}|^2 + \frac{1}{2d}|\widehat{\delta Z_i}|^2 h \le (1+ch)\mathbb{E}_i\left[|\delta Y_{i+1}|^2 + \frac{1}{8d}|\delta Z_{i+1}|^2 h\right] + 2H_i^\theta,$$

where

$$H_i^\theta = (1-\theta)^2 \mathbb{E}_i\big[|\delta f_{i+1}|^2\big]h^2 - \theta^2 \mathbb{E}_i\big[|\delta \widehat{f}_i|^2\big]h^2. \tag{5.2}$$

PROOF. This proof is very similar to that of Proposition 4.9, therefore we omit it. □

We want to control $\mathcal{R}^{\mathcal{S}}(H)$. For the fully implicit scheme ($\theta = 1$), we have $H_i^\theta = -|\delta \widehat{f}_i|^2 h^2 \le 0$ and hence the implicit scheme is stable in the classical sense [of Chassagneux (2012, 2013)] as we have $\mathcal{R}^{\mathcal{S}}(H) \le 0$. The next lemma provides, in our setting, a control on $\mathcal{R}^{\mathcal{S}}(H)$ for any $\theta \ge 1/2$.

LEMMA 5.3. *Let* (HX0), (HY0$_{\text{loc}}$) *hold and take the family* $\{H_i\}_{i=0,\dots,N-1}$ *defined in* (5.2). *Then for* $\theta \ge 1/2$ *there exists a constant* c *such that*

$$\begin{aligned}
\mathcal{R}^{\mathcal{S}}(H) &= \max_{i=0,\dots,N-1} \mathbb{E}\left[\sum_{j=i}^{N-1} e^{c(j-i)h} H_j^\theta\right] \\
&\le c\mathbb{E}\big[|Y_{t_N} - Y_N|^4\big]^{1/2} h^2 + c\mathbb{E}\big[|\bar{Z}_N - Z_N|^2\big]h^2 \\
&\quad + c\left(\sum_{i=0}^{N-1} \tau_i(Y)\right)^{1/2} h + c\left(\sum_{i=0}^{N-1} \tau_i(Z)\right)^{1/2} h.
\end{aligned}$$

PROOF. Let $i \in \{0, \dots, N-1\}$. For $1/2 \le \theta \le 1$, we have $(1-\theta)^2 \le \theta^2$ and, therefore,

$$\begin{aligned}
&\mathbb{E}\Biggl[\sum_{j=i}^{N-1} e^{c(j-i)h} H_j^\theta\Biggr] \\
&\quad\le \theta^2 \mathbb{E}\Biggl[\sum_{j=i}^{N-1} e^{c(j-i)h}\bigl(|\delta f_{j+1}|^2 - |\delta \widehat{f}_j|^2\bigr)h^2\Biggr] \\
&\quad= \theta^2 \mathbb{E}\Biggl[\sum_{j=i}^{N-1} e^{c(j-i)h}\bigl(|\delta f_{j+1}|^2 - |\delta f_j + \beta_j|^2\bigr)h^2\Biggr] \\
&\quad\le \theta^2 \mathbb{E}\Biggl[\sum_{j=i}^{N-1} e^{c(j-i)h}\bigl(e^{ch}|\delta f_{j+1}|^2 - |\delta f_j|^2 - 2\langle \delta f_j, \beta_i\rangle - \beta_j{}^2\bigr)h^2\Biggr] \\
&\quad\le \theta^2 e^{c(N-i)h}\mathbb{E}\bigl[|\delta f_N|^2\bigr]h^2 - 2\theta^2 \sum_{j=i}^{N-1} e^{c(j-i)h}\mathbb{E}\bigl[\langle \delta f_j, \beta_j\rangle\bigr]h^2,
\end{aligned}$$

where $\beta_i := \delta \widehat{f}_j - \delta f_j = f(t_i, X_j, \widehat{Y}_j, \widehat{Z}_j) - f(t_i, X_j, Y_{t_i}, \bar{Z}_{t_i})$ and we used a telescopic sum. Using now ($\mathrm{HY0_{loc}}$) yields

$$\begin{aligned}
\mathbb{E}\bigl[|\delta f_N|^2\bigr] &\le c\mathbb{E}\bigl[1 + |Y_{t_N}|^{4(m-1)} + |Y_N|^{4(m-1)}\bigr]^{1/2}\mathbb{E}\bigl[|Y_{t_N} - Y_N|^4\bigr]^{1/2} \\
&\quad + c\mathbb{E}\bigl[|\bar{Z}_N - Z_N|^2\bigr]
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\bigl[\langle \delta f_i, \beta_i\rangle\bigr]h^2 &\le \mathbb{E}\bigl[|\delta f_i||\beta_i|\bigr]h^2 \\
&\le \mathbb{E}\bigl[\bigl(|\delta f_i| L_y\bigl(1 + |\widehat{Y}_i|^{m-1} + |Y_{t_i}|^{m-1}\bigr)\bigr)^2\bigr]^{1/2}\mathbb{E}\bigl[|\widehat{Y}_i - Y_{t_i}|^2\bigr]^{1/2}h^2 \\
&\quad + \mathbb{E}\bigl[\bigl(L_z|\delta f_i|\bigr)^2\bigr]^{1/2}\mathbb{E}\bigl[|\widehat{Z}_i - \bar{Z}_{t_i}|^2\bigr]^{1/2}h^2 \\
&\le c\mathbb{E}\bigl[B_i^1\bigr]^{1/2}\mathbb{E}\bigl[|\widehat{Y}_i - Y_{t_i}|^2\bigr]^{1/2}h + c\mathbb{E}\bigl[B_i^2\bigr]^{1/2}\mathbb{E}\bigl[|\widehat{Z}_i - \bar{Z}_{t_i}|^2 h\bigr]^{1/2}h,
\end{aligned}$$

where $B_i^2 := |Y_{t_i}|^{2m}h + |Y_i|^{2m}h + |\bar{Z}_{t_i}|^2 h + |Z_i|^2 h$ and

$$B_i^1 := h^2 + |\widehat{Y}_i|^{4m}h^2 + |Y_{t_i}|^{4m}h^2 + |Y_i|^{4m}h^2 + \bigl(|\bar{Z}_{t_i}|^2 h\bigr)^2 + \bigl(|Z_i|^2 h\bigr)^2.$$

From Theorem 2.2, Corollary 3.6, Remark 4.10 and Proposition 5.1, we have for the first term of the above inequality

$$\begin{aligned}
\sum_{i=0}^{N-1}\mathbb{E}\bigl[B_i^1\bigr]^{1/2}\mathbb{E}\bigl[|\widehat{Y}_i - Y_{t_i}|^2\bigr]^{1/2}h &\le \Biggl(\sum_{i=0}^{N-1}\mathbb{E}\bigl[B_i^1\bigr]\Biggr)^{1/2}\Biggl(\sum_{i=0}^{N-1}\tau_i(Y)\Biggr)^{1/2} h \\
&\le c\Biggl(\sum_{i=0}^{N-1}\tau_i(Y)\Biggr)^{1/2} h
\end{aligned}$$

and similarly for the second term

$$\sum_{i=0}^{N-1} \mathbb{E}\big[B_i^2\big]^{1/2}\mathbb{E}\big[|\widehat{Z}_i - \bar{Z}_{t_i}|^2 h\big]^{1/2} h \le c\left(\sum_{i=0}^{N-1} \tau_i(Z)\right)^{1/2} h. \qquad \square$$

5.3. *Convergence of the scheme*. By collecting the above results, we can now prove Theorem 4.2.

PROOF OF THEOREM 4.2. The proof is a combination of the Fundamental Lemmas 4.6 and 4.8, Proposition 4.13 and stability results obtained in this section, namely Proposition 5.2 and Lemma 5.3.

We move to the proof of part (ii), the case $\theta = 1/2$. Since in this case f depends only on y, a quick rerun of arguments of the Fundamental Lemma 4.6, shows there exists a constant $c > 0$ such that

$$\max_{i=0,\dots,N} \mathbb{E}\big[|Y_{t_i} - Y_i|^2\big] \le c\left\{\mathbb{E}\big[|Y_{t_N} - Y_N|^2\big] + \sum_{i=0}^{N-1} \frac{\tau_i(Y)}{h}\right\} + (1+h)\mathcal{R}^{\mathcal{S}}(H).$$

The first two terms on the RHS can be bounded by $ch^{2\gamma} + ch^4$, $c > 0$, using Lemma 4.8 and Proposition 4.13, respectively. By Lemma 5.3, there exists a constant $c > 0$ such that

$$\mathcal{R}^{\mathcal{S}}(H) \le c\mathbb{E}\big[|Y_{t_N} - Y_N|^4\big]^{1/2} h^2 + c\left(\sum_{i=0}^{N-1} \tau_i(Y)\right)^{1/2} h,$$

and using again Lemma 4.8 and Proposition 4.13 yields $\mathcal{R}^{\mathcal{S}}(H) \le ch^{2\gamma+2} + ch^{7/2}$. By joining these results, the theorem's conclusion follows. $\square$

5.4. *Further remarks*. Here, we discuss a true overall second-order scheme, namely a second-order discretization for Z and an intuitive variance reduction technique which we have used throughout but not made formally explicit.

5.4.1. *The candidate for second-order scheme*. For the general case where the driver depends on Z, the approximation for Z_i, namely (4.6), is not enough to obtain a higher order scheme as it is a first-order approximation. The proper higher order scheme in its full generality follows by applying the trapezoidal rule to *all* integrals present in (4.14); as is done for (4.13). With some manipulation (left to the reader), we end up with the following approximation for Z_i [compare with (4.6)]:

$$Z_i = \frac{2}{h}\mathbb{E}_i\big[\Delta W_{i+1}\big(Y_{t_{i+1}} + (1-\theta) f(t_i, X_{i+1}, Y_{i+1}, Z_{i+1})h\big)\big] - \mathbb{E}_i[Z_{i+1}],$$

with $\theta = 1/2$, the terminal condition $Y_N = g(X_N)$, along with (4.5) and a suitable approximation for Z_T. An approximation for Z_T is not trivial and could, for instance, be found via Malliavin calculus. The general treatment of such a scheme is left for future research.

Another type of second-order scheme can be found in Crisan and Manolarakis (2010); the approximation there is based on Itô–Taylor expansions.

5.4.2. *Controlling the variance of the scheme*. If we use the notation set up in Section 4.5, the approximation (4.6) can be written out as $Z_i = \mathbb{E}_i[\Delta W_{i+1} A_{i+1}]/h$. We point out that implementation-wise it is better to use the lower variance approximation (4.20) instead of (4.6), that is, to use

$$Z_i = \frac{1}{h}\mathbb{E}_i\big[\Delta W_{i+1}\big(A_{i+1} - \mathbb{E}_i[A_{i+1}]\big)\big], \qquad i = 0, \dots, N-1.$$

This does not lead to a relevant additional computation effort, as $\mathbb{E}_i[A_{i+1}]$ must be computed for the estimation of the Y_i component. To avoid a long analysis, we make some simplifying assumptions in order to better explain the gain: assume $X_t = x + W_t$ and that we are about to compute Z_0 (a standard expectation); assume further (via Doob–Dynkin lemma) that A_1 can be written as[12] $A_1 = \varphi(X_1) = \varphi(x + \Delta W_1)$ where φ has some regularity so that

$$\varphi(x + \Delta W_1) = \varphi(x) + \varphi'(x)(\Delta W_1) + \tfrac{1}{2}\varphi''(x^*)(\Delta W_1)^2,$$

where x^* lies between x and $x + \Delta W_1$. Then the Monte Carlo (MC) estimator for Z_0 from (4.6), with M samples of the normal $\mathcal{N}(0,1)$ distribution given by $\{\mathcal{N}^\lambda\}_{\lambda=1,\dots,M}$, and its standard deviation (St.d.) are

$$Z_0^{\mathrm{MC},(4.6)} = \frac{1}{M}\sum_{\lambda=1}^{M} \frac{\sqrt{h}\mathcal{N}^\lambda}{h}\varphi\big(x + \sqrt{h}\mathcal{N}^\lambda\big) \qquad \text{with St.d.} \approx \frac{|\varphi(x)|}{\sqrt{h}\sqrt{M}}.$$

Using (4.20) instead of (4.6) to compute Z_0 would produce the MC estimator and its St.d.

$$Z_0^{\mathrm{MC},(4.20)} = \frac{1}{M}\sum_{\lambda=1}^{M} \frac{\sqrt{h}\mathcal{N}^\lambda}{h}\big(\varphi\big(x + \sqrt{h}\mathcal{N}^\lambda\big) - \varphi(x)\big) \qquad \text{with St.d.} \approx \frac{|\varphi'(x)|}{\sqrt{M}}.$$

Compare now the standard deviation of both estimators. It is crucial for the stability that the denominator of the variance of $Z_0^{\mathrm{MC},(4.20)}$ lacks that $\sqrt{h}$ term. If M is kept fixed then as h gets smaller we expect $Z_0^{\mathrm{MC},(4.6)}$ to blow up while $Z_0^{\mathrm{MC},(4.20)}$ will remain controlled (assuming φ can be controlled[13]). This can be numerically confirmed in Alanko and Avellaneda (2013).

We point out that this simple trick can be adapted to the scheme proposed in the next section as well as to the computation of the second-order scheme proposed previously.

[12] If the reader is aware of how conditional expectations in the BSDE framework are calculated, say, for example, via projection over a basis of functions, having a function φ is expected.

[13] In Gobet and Turkedjiev (2011), it is shown for the locally Lipschitz driver case that φ is indeed a Lipschitz function of its variables.

6. Convergence of the tamed explicit scheme. We now turn our attention back to the explicit scheme. Unlike the case $\theta \in [1/2, 1]$, when $\theta < 1/2$, the local estimates of Proposition 4.9 cannot be extended to the global ones (as in Proposition 5.1). Consequently, we also do not have a control over the stability remainder $\mathcal{R}^{\mathcal{S}}(H)$ (see Definition 4.4). In fact, as the motivating example of the Introduction shows, the scheme can explode. To remedy to this, we consider the tamed explicit scheme, described in (4.7)–(4.8), which in turn corresponds to a truncation procedure applied to the original BSDE, and show that this scheme converges. Our analysis yields as a by-product sufficient conditions under which the naive explicit scheme converges (see Remark 6.6).

REMARK 6.1 ($m > 1$). In this section, we focus exclusively on the case $m > 1$ in assumption (HY0). The easier case $m = 1$ does not require taming and stability of the scheme results from a straightforward adaptation of the proof of Proposition 6.4.

6.1. *Principle.* The idea is that with the truncation functions T_{L_h} and T_{K_h} [recall the scheme (4.7)–(4.8)], one cannot only obtain uniform integrability bounds for the scheme, but also a pathwise bound, ensuring that the output $\{Y_i\}_{i=0,\dots,N}$ stays under a certain threshold, under which the scheme is found to be stable in the sense of (4.9) with $H_i = 0$.

Note that this tamed scheme is not exactly the scheme (4.5)–(4.6) with $\theta = 0$. However, it can be seen as the case $\theta = 0$ with the functions $T_{L_h} \circ g$ and $f(\cdot, T_{K_h}(\cdot), \cdot, \cdot)$ instead of g and f. They satisfy the same properties with the same constants, so we can reuse the results of Section 4.

Because the scheme is controlled, we naturally compare first its output $\{(Y_i, Z_i)\}_{i\in\{0,\dots,N\}}$ to $(Y'_{t_i}, \bar{Z}'_{t_i})_{t_i\in\pi}$, where $(Y'_t, Z'_t)_{t\in[0,T]}$ is the solution to the BSDE (1.2) with controlled coefficients, for $t \in [0, T]$

$$Y'_t = T_{L_h}\big(g(X_T)\big) + \int_t^T f\big(u, T_{K_h}(X_u), Y'_u, Z'_u\big)\,\mathrm{d}u - \int_t^T Z'_u\,\mathrm{d}W_u. \tag{6.1}$$

This part of analysis follows the methodology used above.

In a second step, it is enough to estimate the distance between the solution (Y', Z') of the truncated BSDE (6.1) and the solution (Y, Z) of the original BSDE (1.2) in order to conclude to the convergence of the scheme.

In line with Sections 4 and 5, we define $\{\bar{Z}'_{t_i}\}_{t_i\in\pi}$ as in (3.23), $\widehat{Y}_i = Y_{i,(Y'_{i+1},\bar{Z}'_{i+1})}$ and $\widehat{Z}_i = Z_{i,(Y'_{i+1},\bar{Z}'_{i+1})}$ for $i = 0, \dots, N-1$, more precisely

$$\widehat{Y}_i := \mathbb{E}_i\big[Y'_{t_{i+1}} + f\big(t_{i+1}, T_{K_h}(X_{i+1}), Y'_{t_{i+1}}, \bar{Z}'_{t_{i+1}}\big)h\big], \tag{6.2}$$

$$\widehat{Z}_i := \mathbb{E}_i\bigg[\frac{\Delta W_{i+1}}{h}\big(Y'_{t_{i+1}} + f_h\big(t_{i+1}, X_{i+1}, Y'_{t_{i+1}}, \bar{Z}'_{t_{i+1}}\big)h\big)\bigg]. \tag{6.3}$$

6.2. *Integrability for the scheme*. We now show that the tamed Euler scheme has the property that $|Y_i| \le h^{-1/(2m-2)}$ for all $i \in \{0,\dots,N\}$. This is already true for $Y_N = T_{L_h}(g(X_N))$ by construction. In the next two propositions, we will show that this bound propagates through time.

PROPOSITION 6.2. *Assume* (HX0), (HY0) *and that* $h \le 1/(32dL_z^2)$. *If for a given* $i \in \{0,\dots,N-1\}$ *one has* $|Y_{i+1}| \le h^{-1/(2m-2)}$, *then one also has*

$$|Y_i|^2 + \frac{1}{d}|Z_i|^2 h \le (1+c_1 h)\mathbb{E}_i\left[|Y_{i+1}|^2 + \frac{1}{4d}|Z_{i+1}|^2 h\right] + c_2 h + c_2 h \mathbb{E}_i\left[\left|T_{K_h}(X_{i+1})\right|^2\right].$$

PROOF. Take $i \in \{0,\dots,N-1\}$. We have seen in the proof of Proposition 4.9, equation (4.23) that, since $\theta = 0$,

$$|Y_i|^2 + \frac{1}{d}|Z_i|^2 h \le \left(1 + 2(L_y+\alpha')h\right)\mathbb{E}_i\left[|Y_{i+1}|^2\right] + \frac{3L_z^2}{2\alpha'}\mathbb{E}_i\left[|Z_{i+1}|^2\right]h + \mathbb{E}_i\left[B(i+1,\alpha')\right] + H_i^0,$$

where $B(i+1,\alpha') := (3L^2 h + 3L_x^2|T_{K_h}(X_{i+1})|^2 h)/2\alpha'$ and

$$H_i^0 = \mathbb{E}_i\left[|f_{i+1}|^2\right]h^2 = \mathbb{E}_i\left[\left|f\left(t_{i+1}, T_{K_N}(X_{i+1}), Y_{i+1}, Z_{i+1}\right)\right|^2\right]h^2.$$

Using (HY0) and the fact that $|Y_{i+1}|^{2(m-1)}h \le 1$, we have

$$\begin{aligned}|f_{i+1}|^2 h^2 &\le 4L^2h^2 + 4L_x^2\left|T_{K_h}(X_{i+1})\right|^2h^2 \\ &\quad + 4L_y^2\left[|Y_{i+1}|^{2(m-1)}h\right]|Y_{i+1}|^2 h + 4L_z^2|Z_{i+1}|^2h^2 \\ &\le 4L^2h^2 + 4L_x^2\left|T_{K_h}(X_{i+1})\right|^2h^2 + 4L_y^2|Y_{i+1}|^2h + 4L_z^2 h|Z_{i+1}|^2 h,\end{aligned}$$

so we have in the end

$$\begin{aligned}|Y_i|^2 + \frac{1}{d}|Z_i|^2 h &\le \left(1 + 2\left(L_y + \alpha' + 2L_y^2\right)h\right)\mathbb{E}_i\left[|Y_{i+1}|^2\right] \\ &\quad + \left(\frac{3L_z^2}{2\alpha'} + 4L_z^2 h\right)\mathbb{E}_i\left[|Z_{i+1}|^2\right]h \\ &\quad + \left(\frac{3L^2}{2\alpha'} + 4L^2h\right)h + \left(\frac{3L_x^2}{2\alpha'} + 4L_x^2h\right)\mathbb{E}_i\left[\left|T_{K_h}(X_{i+1})\right|^2\right]h.\end{aligned}$$

Choose now $\alpha' = 12dL_z^2$ [so that $3L_z^2/(2\alpha') \le 1/(8d)$] and combine with the restriction $h \le 1/(32dL_z^2)$ (so that $4L_z^2 h \le \frac{1}{8d}$). Taking $c_1 = 2(L_y + 12dL_z^2 + 2L_y^2)$ and

$$c_2 = \max\left\{\frac{3L^2}{24dL_z^2} + \frac{4L^2}{32dL_z^2}, \frac{3L_x^2}{24dL_z^2} + \frac{4L_x^2}{32dL_z^2}\right\} = \max\left\{\frac{L^2}{4dL_z^2}, \frac{L_x^2}{4dL_z^2}\right\},$$

and noting that $1/(4d) \le (1+c_1h)/(4d)$, we find the required estimate

$$|Y_i|^2 + \frac{1}{d}|Z_i|^2 h \le (1+c_1h)\mathbb{E}_i\left[|Y_{i+1}|^2 + \frac{1}{4d}|Z_{i+1}|^2 h\right] + c_2h + c_2h\mathbb{E}_i\left[\left|T_{K_h}(X_{i+1})\right|^2\right]. \qquad \square$$

We can then use this local bound to obtain the following pathwise bound.

PROPOSITION 6.3. *Let* (HX0) *and* (HY0) *hold. For any* $i \in \{0,\dots,N-1\}$,

$$|Y_i|^2 + \frac{1}{4d}|Z_i|^2 h + \frac{3}{4d}\mathbb{E}_i\left[\sum_{j=i}^{N-1}|Z_j|^2 h\right]$$
$$\le e^{c_1(N-i)h}\mathbb{E}_i\left[|Y_N|^2\right] + e^{c_1(N-1-i)h}\left(\sum_{j=i}^{N-1} c_2h + c_2h\mathbb{E}_i\left[\left|T_{K_h}(X_{i+1})\right|^2\right]\right).$$

This implies in particular that $|Y_i| \le h^{-1/(2m-2)}$.

PROOF. The proof goes by induction. The case $i = N$ is clear. If the estimate is true for $i+1$, noting that $|Y_N| \le L_h$, $|T_{K_h}(x)| \le K_h$ and $e^{c_1T}(L_h^2 + c_2T + c_2TK_h^2) \le h^{-1/(m-1)}$, we see that $|Y_{i+1}|^2 \le h^{-1/(m-1)}$. Then, combining the estimate of Proposition 6.2 and the estimate for $i+1$ (from the induction assumption), in the same way as in Lemma A.4, we obtain the desired estimate for i. $\square$

In view of the previous bound, we can derive a similar estimate for the solution (Y', Z') to (6.1). Namely, using (2.5) with $\alpha = 12dL_z^2$ and combining it further with (HY0), we have

$$\begin{aligned}
\left|Y'_t\right|^2 &\le e^{2(L_y+12dL_z^2)(T-t)} \\
&\quad\times \mathbb{E}_t\left[\left|T_{L_h}\big(g(X_T)\big)\right|^2 + \int_t^T \frac{1}{16dL_z^2}\left|f\big(u, T_{K_h}(X_u), 0, 0\big)\right|^2 \mathrm{d}u\right] \\
&\le e^{c_1(T-t)}\mathbb{E}_t\left[\left|T_{L_h}\big(g(X_T)\big)\right|^2 + \int_t^T \frac{1}{8dL_z^2}\left(L^2 + L_x^2|T_{K_h}(X_u)|^2\right)\mathrm{d}u\right] \\
&\le e^{c_1T}\left(L_h^2 + c_2T + c_2TK_h^2\right) \\
&\le \left(\frac{1}{h}\right)^{1/(m-1)},
\end{aligned}$$

implying in particular that $|Y'_{t_i}| \le h^{-1/(2m-2)}$ for all i.

These two estimates, ensuring that both Y_i and Y'_{t_i} are bounded by $h^{-1/(2m-2)}$ will be useful in the analysis of the global error, since the explicit scheme is found to be stable under this threshold.

6.3. *Stability of the scheme.* As previously, for any $i \in \{0, \dots, N-1\}$ we use the notation $\delta Y_{i+1} := Y'_{t_{i+1}} - Y_{i+1}$ and $\delta Z_{i+1} := \bar{Z}'_{t_{i+1}} - Z_{i+1}$, as well as $\delta A_{i+1} := \delta Y_{i+1} + \delta f_{i+1} h$ where δf_{i+1} is given by

$$\delta f_{i+1} := f\big(t_{i+1}, T_{K_h}(X_{i+1}), Y'_{t_{i+1}}, \bar{Z}'_{i+1}\big) - f\big(t_{i+1}, T_{K_h}(X_{i+1}), Y_{i+1}, Z_{i+1}\big).$$

Then, denoting $\widehat{\delta Y}_i = \widehat{Y}_i - Y_i$ and $\widehat{\delta Z}_i = \widehat{Z}_i - Z_i$, we can write

$$\widehat{\delta Y}_i = \mathbb{E}_i[\delta A_{i+1}] \quad \text{and} \quad \widehat{\delta Z}_i = \mathbb{E}_i\left[\frac{1}{h}\Delta W_{i+1}\delta A_{i+1}\right].$$

We now proceed to show that, because the two inputs satisfy $|Y_{i+1}|, |Y'_{t_{i+1}}| \le h^{-1/(2m-2)}$, the scheme is stable in the sense that we can obtain the estimate (4.9) with $H_i = 0$.

PROPOSITION 6.4. *Assume* (HX0) *and* ($\mathrm{HY0_{loc}}$). *Then there exists a constant* c *for any* $h \le \min\{1, 1/32dL_z^2\}$, *such that for* $i \in \{0, \dots, N-1\}$

$$|\widehat{\delta Y}_i|^2 + \frac{1}{d}|\widehat{\delta Z}_i|^2 h \le (1+ch)\mathbb{E}_i\left[|\delta Y_{i+1}|^2 + \frac{1}{4d}|\delta Z_{i+1}|^2 h\right]. \tag{6.4}$$

PROOF. Let $i \in \{0, \dots, N-1\}$. Just like for Proposition 5.2, the proof mimics the computations of the proof of Proposition 4.9 with only a small adjustment for the constants. However, a different argumentation for the term $H_i^0 = |\delta f_{i+1}|^2 h^2$ is required. Using ($\mathrm{HY0_{loc}}$), $h \le 1$ and the bounds $|Y'_{t_{i+1}}|^{2(m-1)} h$, $|Y'_{t_{i+1}}|^{2(m-1)} h \le 1$, we have

$$\begin{aligned}
|\delta f_{i+1}|^2 h^2 &\le 2L_y^2\big(1 + |Y'_{t_{i+1}}|^{2(m-1)} + |Y_{i+1}|^{2(m-1)}\big)\big|Y'_{t_{i+1}} - Y_{i+1}\big|^2 h^2 \\
&\quad + 2L_z^2\big|\bar{Z}'_{t_{i+1}} - Z_{i+1}\big|^2 h^2 \\
&= 2L_y^2\big(h + |Y'_{t_{i+1}}|^{2(m-1)} h + |Y_{i+1}|^{2(m-1)} h\big) h \big|Y'_{t_{i+1}} - Y_{i+1}\big|^2 \\
&\quad + 2L_z^2 h\big|\bar{Z}'_{t_{i+1}} - Z_{i+1}\big|^2 h \\
&\le 6L_y^2 h|\delta Y_{i+1}|^2 + 2L_z^2 h|\delta Z_{i+1}|^2 h.
\end{aligned}$$

The rest follows as in the proof of Proposition 4.9. □

6.4. *Convergence of the scheme.* The convergence of the scheme is achieved by controlling both the (squared) error committed by the truncation procedure, $\|Y - Y'\|^2_{\mathcal{S}^2} + \|Z - Z'\|^2_{\mathcal{H}^2}$, as a function of the time step, and by controlling the numerical approximation (4.7)–(4.8) of the solution (Y', Z') to (6.1).

Distance between $(Y_i, Z_i)_i$ *and* $(Y'_{t_i}, \bar{Z}'_{t_i})_i$. We estimate this distance by using the Fundamental Lemma 4.6.

The tamed scheme (4.7)–(4.8) is the $\theta = 0$ scheme (4.5)–(4.6) with the coefficient $f(\cdot, \cdot, T_{K_h}(\cdot), \cdot)$ and terminal condition $T_{L_h} \circ g$ having the same Lipschitz constant as f and g. So the results of Section 4 apply. In particular, Lemma 4.8 controls the error on the terminal condition.

Similarly, Lemmas 4.11 and 4.12 are still valid with the same constants. The only difference is that the path-regularity involved is now that of (Y', Z'), but since $T_{L_h} \circ g$ is still Lipschitz, Theorem 3.5 indeed applies to (Y', Z'). So Proposition 4.13 applies, to control the sum of the one-step discretization errors.

Finally, we have just proven with Proposition 6.4 that the scheme is stable with $H_i^0 = 0$, so $\mathcal{R}^{\mathcal{S}}(H) = 0$. We can therefore conclude via Lemma 4.6 that

$$
\begin{aligned}
&\max_{i=0,\dots,N} \mathbb{E}[|Y'_{t_i} - Y_i|^2] + \sum_{i=0}^{N-1} \mathbb{E}[|\bar{Z}'_{t_i} - Z_i|^2]h \\
&\quad \le c(\mathbb{E}[|Y'_{t_N} - Y_N|^2] + \mathbb{E}[|\bar{Z}'_{t_N} - Z_N|^2]h) \\
&\qquad + c \sum_{i=0}^{N-1} \left(\frac{1}{h}\tau_i(Y) + \tau_i(Z)\right) + 0 \\
&\quad \le ch.
\end{aligned}
\tag{6.5}
$$

We remark that the thresholds L_h and K_h have no effect in this estimation.

The distance between $(Y'_{t_i}, \bar{Z}'_{t_i})_i$ *and* $(Y_{t_i}, \bar{Z}_{t_i})_i$. We now estimate the distance between $(Y'_{t_i}, \bar{Z}'_{t_i})_i$ and $(Y_{t_i}, \bar{Z}_{t_i})_i$, that is, between (6.1) and (1.2), which gathers all the error induced by the taming. In order to estimate this error, we need to have an estimation of the L^2-distance between X_u and $T_{K_h}(X_u)$ on the one hand, and $g(X_T)$ and $T_{L_h}(g(X_T))$ on the other. We give a general estimation for this below.

PROPOSITION 6.5. *Let ξ be a random variable in L^q for some $q > 2$, and $L > 0$. Then we have*

$$
\mathbb{E}[|\xi - T_L(\xi)|^2] \le 4\mathbb{E}[|\xi|^q]\left(\frac{1}{L}\right)^{q-2}.
$$

PROOF. Using the facts that $T_L(x) = x$ for $|x| \le L$ and that $|T_L(\xi)| \le |\xi|$, together with the Hölder and the Markov inequalities, we have

$$
\begin{aligned}
\mathbb{E}[|\xi - T_L(\xi)|^2] &= \mathbb{E}[|\xi - T_L(\xi)|^2 \mathbb{1}_{\{|\xi| \ge L\}}] \\
&\le 4\mathbb{E}[|\xi|^2 \mathbb{1}_{\{|\xi| \ge L\}}] \\
&\le 4\mathbb{E}[|\xi|^q]^{2/q} \mathbb{P}[|\xi| \ge L]^{1-2/q}
\end{aligned}
$$

$$\leq 4\mathbb{E}[|\xi|^q]^{2/q}\left(\frac{\mathbb{E}[|\xi|^q]}{L^q}\right)^{1-2/q}$$
$$= 4\mathbb{E}[|\xi|^q]\left(\frac{1}{L}\right)^{q(1-2/q)}. \qquad \square$$

Now, via Jensen's inequality we have

$$\begin{aligned}|\bar{Z}_{t_i} - \bar{Z}'_{t_i}|^2 h &= \left|\frac{1}{h}\mathbb{E}_i\left[\int_{t_i}^{t_{i+1}} Z_u \,\mathrm{d}u\right] - \frac{1}{h}\mathbb{E}_i\left[\int_{t_i}^{t_{i+1}} Z'_u \,\mathrm{d}u\right]\right|^2 h \\ &\leq \mathbb{E}_i\left[\int_{t_i}^{t_{i+1}} |Z_u - Z'_u|^2 \,\mathrm{d}u\right],\end{aligned}$$

from which it clearly follows that

$$\begin{aligned}&\max_{i=0,\dots,N} \mathbb{E}[|Y_{t_i} - Y'_{t_i}|^2] + \sum_{i=0}^{N-1} \mathbb{E}[|\bar{Z}_{t_i} - \bar{Z}'_{t_i}|^2]h \\ &\quad \leq \sup_{t\in[0,T]} \mathbb{E}[|Y_t - Y'_t|^2] + \mathbb{E}\left[\int_0^T |Z_u - Z'_u|^2 \,\mathrm{d}u\right].\end{aligned}$$

From the a priori estimate (2.6), we have

$$\begin{aligned}&\sup_{t\in[0,T]} \mathbb{E}[|Y_t - Y'_t|^2] + \mathbb{E}\left[\int_0^T |Z_u - Z'_u|^2 \,\mathrm{d}u\right] \\ &\quad \leq c\Big(\mathbb{E}[|g(X_T) - T_{L_h}(g(X_T))|^2] \\ &\qquad + \mathbb{E}\left[\int_0^T |f(u, X_u, Y'_u, Z'_u) - f(u, T_{K_h}(X_u), Y'_u, Z'_u)|^2 \,\mathrm{d}u\right]\Big) \\ &\quad \leq c\left(\mathbb{E}[|g(X_T) - T_{L_h}(g(X_T))|^2] + L_x^2 \int_0^T \mathbb{E}[|X_u - T_{K_h}(X_u)|^2]\,\mathrm{d}u\right) \\ &\quad \leq c\left(4\left(\frac{1}{L_h}\right)^{2m-2}\mathbb{E}[|g(X_T)|^{2m}] + \left(\frac{1}{K_h}\right)^{2m-2} 4L_x^2 \int_0^T \mathbb{E}[|X_u|^{2m}]\,\mathrm{d}u\right),\end{aligned}$$

thanks to Proposition 6.5. Now, since $X \in \mathcal{S}^{2m}$ (Theorem 2.2), g is of linear growth, and L_h and K_h are of order $h^{-1/(2m-2)}$, we can conclude that

$$\max_{i=0,\dots,N} \mathbb{E}[|Y_{t_i} - Y'_{t_i}|^2] + \sum_{i=0}^{N-1} \mathbb{E}[|\bar{Z}_{t_i} - \bar{Z}'_{t_i}|^2]h \leq ch. \tag{6.6}$$

The proof of the Theorem 4.3. By collecting the above results, we can now prove Theorem 4.3.

PROOF OF THEOREM 4.3. To prove this theorem, that is, that $\mathrm{ERR}_\pi(Y, Z) \le ch^{1/2}$ [see (4.4)], we use the triangular inequality and dominate $\mathrm{ERR}_\pi(Y, Z)$ by the sum of: (i) the distance between the solution (Y, Z) to the original BSDE (1.2) and the solution (Y', Z') to the truncated BSDE (6.1), and (ii) the distance between $(Y'_{t_i}, \bar{Z}'_{t_i})_{t_i\in\pi}$ and the $\{(Y_i, Z_i)\}_{i\in\{0,\dots,N\}}$ [from the scheme (4.7)–(4.8)]. The estimate for the first difference is given by (6.6). The estimate for the second is given by (6.5). Hence, the result. □

REMARK 6.6. We see from the proofs of Propositions 6.2 and 6.3 that if $x \mapsto f(t, x, y, z)$ is bounded (say, by K) uniformly in the other variables and the terminal condition g is bounded, then the naive explicit scheme [i.e., (4.5)–(4.6) with $\theta = 0$] converges. Under these conditions, it is suitable to use the explicit backward Euler scheme.

7. Numerical experiments. We conclude with some numerical experiments for the convergence of the introduced schemes. In this work, we are concerned only with the time-discretization, but in order to implement a scheme, we need to further approximate the required conditional expectations. For this, we use the method of regression on a basis functions as in Gobet, Lemor and Warin (2005), Gobet and Turkedjiev (2011). Following Gobet, Lemor and Warin (2005), we work with (Hermite) polynomials up to a certain degree K. Here, we do not aim at studying the effect of the number K of basis functions or the number M of diffusion paths $\{X_i^m\}_{i=0,\dots,N}^{m=1,\dots,M}$. Rather, we choose K and M big enough so that (a) the variance of the results is small enough, and (b) the effect of approximating the conditional expectation is negligible and so what we measure is indeed the effect on the time-discretization of the time-step $h = T/N$.

In all the examples below, we fix terminal time T and want to compute an approximation of $u(t, X_t) = Y_t =: Y_t^{\mathrm{true}}$. Since in this section we use grids with different numbers N of intervals, we do not omit the superscripts and denote by Y_i^N the scheme's approximation of $Y_{t_i}^{\mathrm{true}}$. When the explicit solution to the FBSDE is known, we can measure the error of the numerical approximation by estimating $\mathrm{ERR}(Y^N) = \max_i \mathbb{E}[|Y_{t_i}^{\mathrm{true}} - Y_i^N|^2]^{1/2}$. When the explicit solution is not known, we can compute

$$e(N) := \max_{i=0,\dots,N} \mathbb{E}[|Y_i^N - Y_{2i}^{2N}|^2]^{1/2}. \tag{7.1}$$

By observing the convergence of $e(N)$ we can measure the convergence rate of the scheme even when we do not know the true solution. Indeed, assume that for constants c and γ, for any N and any $i = 0, \dots, N$ we have

$$\mathbb{E}[|Y_i^N - Y_{2i}^{2N}|^2]^{1/2} \le cN^{-\gamma},$$

$$\Longrightarrow \quad \mathbb{E}[|Y_i^N - Y_{t_i}^{\mathrm{true}}|^2]^{1/2} \le \sum_{k=0}^{\infty} c(2^k N)^{-\gamma} = \frac{cN^{-\gamma}}{1-(1/2)^\gamma} = c'N^{-\gamma},$$

given that the scheme converges.

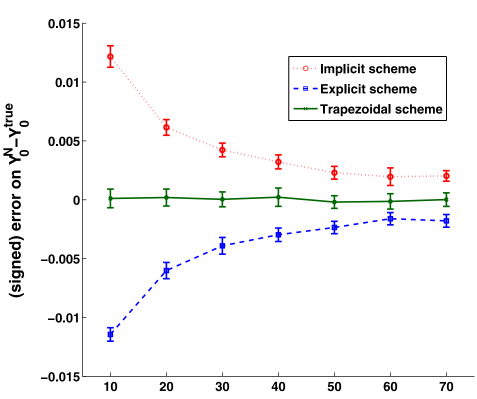
We computed the approximation processes (Y_i^N) and (Y_i^{2N}) using the same sample of Brownian increments. For each measurement, we launched the scheme 10 times and averaged the results.

Example 1—*Numerical approximation for Example* 2.8. We consider the motivating FitzHugh–Nagumo PDE and the terminal condition g of Example 2.8 with $a = -1$, for which $f(t, x, y, z) = -y^3 + y$: a cubic polynomial (without quadratic terms). To solve the implicit equation [see (4.5)], we can use Cardano's formula to compute the single real root of the polynomial equation.

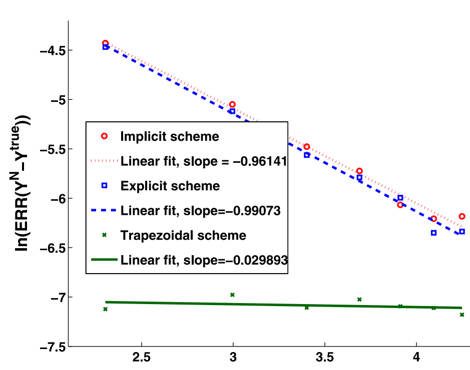
We take $T = 1$ and $x_0 = 3/2$. The solution to the PDE is given by (2.11). We compute the error for various values of N, and this for the explicit scheme ($\theta = 0$, which converges in that case since g is bounded—see Remark 6.6), the implicit scheme ($\theta = 1$) and the trapezoidal scheme [$\theta = 1/2$, note that we are under the extra assumptions made in Theorem 4.2(ii)].

In Figure 1(a), we see that the implicit scheme overshoots the true solution while the explicit one undershoots it; the trapezoidal scheme performs better in any grid. The convergence rates, as measured using $\mathrm{ERR}(Y^N)$, are presented in Figure 1(b). For the trapezoidal scheme, the error for any N is very small and the variance of the results is not negligible, hence we are not able to measure the convergence rate as accurately. The experimental rate seems to be lower than that of the explicit and implicit; see Table 1. We note, however, that the error is already much lower than those in the other schemes.

Both the implicit and explicit schemes are found to converge with rate 1. This does not mean that the estimates in Theorems 4.2 and 4.3 (or that in the Fundamen-



(a)



(b)

FIG. 1. (a) *Differences* $Y_0^N - Y_0^{\mathrm{true}}$ *for each scheme as functions on the number* N *of time intervals*. (b) *Convergence rates obtained via linear fits on the log–log plots of* $\mathrm{ERR}(Y^N)$. *We used* $N \in \{10, 20, 30, 40, 50, 60, 70\}$, *Hermite polynomials up to degree* $K = 7$, $M = 2 \times 10^5$ *and* 10 *simulations for each point*.

TABLE 1
Estimated rates (value of slope) for the experiment reported in Figure 1

Scheme	**Rate via ERR(Y^N)**	**Rate via $e(N)$**
Implicit	−0.96141	−1.00460
Explicit	−0.99073	−0.98372
Trapezoidal	−0.02989	−0.33775

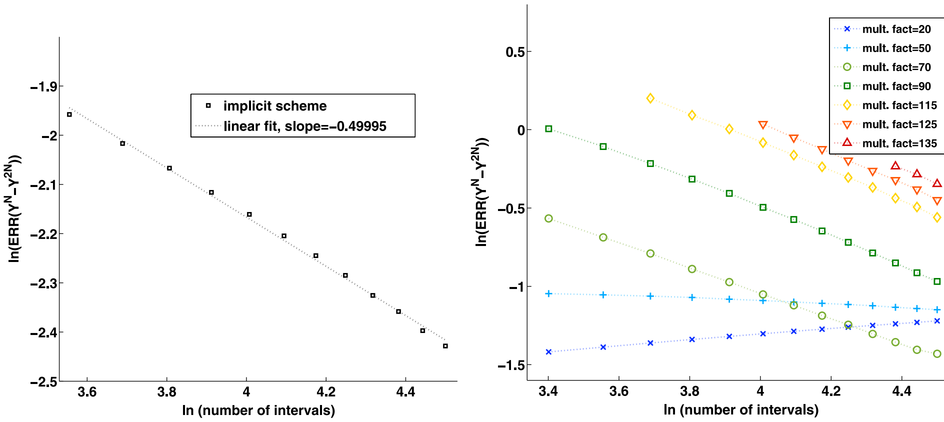
tal Lemma 4.6) are too conservative in all generality, but is simply due to the particularity of the equation studied. On the one hand, the estimates of Theorems 4.2 and 4.3 rely on the estimate of Proposition 4.13 (on the local discretization errors) and so on the regularity of b, σ, f and g. We worked under the minimal assumption (HY0$_{\text{loc}}$) assuming no differentiability. Nonetheless, in this example all involved functions are smooth (leading to a smooth solution u to the PDE) and so this term ends up converging faster (see also Remark 4.14). On the other hand, the estimates of Theorems 4.2 and 4.3 also rely on the estimate of Lemma 4.8 (on the terminal condition error) which again holds under the mere assumption (HX0) for b and σ. But here (X_t) is the Brownian motion an its approximation (X_i^N) is exact, instead of being only of order $\gamma = 1/2$ in the case of Euler–Maruyama scheme.

As we could verify in our simulations, the computational time is the same for all the schemes with $\theta > 0$, as expected. On the other hand, similar to the case of ODEs and SDEs, the convergence rate for $\theta \in]1/2, 1[$ is no better than for $\theta = 1$. However, the latter choice is more stable [compare with the definition of $\mathcal{R}^{\mathcal{S}}(H)$ and H_i^θ] while $\theta = 1/2$ provides the smallest error. A more detailed comparison between the different implicit-dominating schemes is left to a forthcoming work.

Finally, while we were able to compute ERR(Y^N) in this example, we also computed $e(N)$. Since we approximated the solution using polynomials up to degree K, the full (implemented) scheme computes in fact an approximated process $Y^{N,K}$. As $N \to +\infty$, this does not strictly converge to Y^{true} but rather to some Y^K. The convergence of $e(N)$ therefore better captures the convergence of $Y^{N,K}$ to its limit, Y^K and, therefore, yields slightly different rates.

Example 2—*Unbounded terminal condition*. To emphasize the contribution of this work, we analyze in more detail the unbounded terminal condition case for which one needs to take either the implicit scheme or the explicit scheme with truncated terminal condition. More precisely, we take $g(x) = x$, together with the driver $f(y) = -y^3$. For the forward process, we take the geometric Brownian motion with $b(x) = x/2$ and $\sigma(x) = x/2$, started at $x_0 = 2$. We choose $T = 1$.

Figure 2(a) shows the convergence of $e(N)$ [see (7.1)] for the implicit scheme, while Figure 2(b) shows the same computations for the truncated explicit scheme. The implicit scheme converges with the rate $1/2$, as expected. Concerning the



(a) (b)

FIG. 2. (a) *Convergence of* $e(N)$ *for the implicit scheme.* (b) *Convergence of* $e(N)$ *for the tamed explicit scheme and various values of the multiplying factor. In both cases, we used* $N \in \{5i : i = 7, \dots, 18\}$, $K = 4$, $M = 10^5$ *and* 10 *simulations for each point. The results are plotted in log–log scale.*

truncated explicit scheme [Figure 2(b)] we observed through several trials that its behavior is quite sensitive to the truncation level L_h (defined in Section 4.2.2).[14] Our asymptotic, theoretical results [see (4.7), (4.8) and Theorem 4.3] suggest taking for this particular example L_h as

$$L_h = \frac{1}{\sqrt{3}} e^{-(1/2)6T} \left(\frac{1}{h}\right)^{1/4}.$$

We found, however, that this seems to be too conservative for practical simulations. To better understand the impact of truncation, we introduced a multiplying factor $\alpha > 0$ and truncate at the level αL_h instead of L_h. In Figure 2(b) and Table 2, we sum up our findings. In Table 2, one sees the various multiplying factors and the corresponding estimated rates [for the sequence $e(N)$ defined in (7.1)].

By looking at Figure 2(b), we see that the situation is complex and a separate argumentation is required for "small" and "big" multiplying factors. For

TABLE 2
Estimated rate for the truncated explicit scheme at truncation level αL_h

Mult. factor α	**20**	**50**	**70**	**90**	**115**	**125**	**135**
Rate	0.179	−0.096	−0.801	−0.896	−0.929	−0.970	−0.955

[14]This echoes the findings of Chassagneux and Richou (2013).

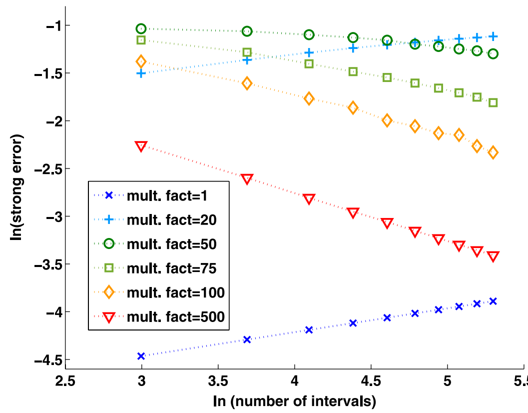


FIG. 3. *Convergence of the error* $\mathbb{E}[|T_{L_{1/N}}(g(X_N^N)) - T_{L_{1/(2N)}}(g(X_{2N}^{2N}))|^2]^{1/2}$ *on the terminal condition, computed for* $N \in \{20i : i = 1, \dots, 10\}$. *Plot in log–log scale with different levels of truncation* $L_{1/N} = \alpha L_h$, *done with* $M = 10^5$ *and* 10 *simulations for each point. The estimated slopes are, for the corresponding multiplicative factors*: 0.25, 0.17, −0.12, −0.29, −0.41, −0.50 (*reading the legend from top to bottom*).

α too small (up to 40), the scheme does not seem to converge. This is due to the fact that a significant number of forward paths fall beyond truncation levels $\alpha L_{1/N}$ and $\alpha L_{1/(2N)}$. Consequently, the strong convergence property for the forward approximation does not guarantee that the quantity $\mathbb{E}[|T_{L_{1/N}}(g(X_T^N)) - T_{L_{1/(2N)}}(g(X_T^{2N}))|^2]^{1/2}$ decays with the rate 1/2, as is shown in Figure 3. This lack of "good convergence" at the terminal time then translates into a deterioration of the convergence rate for the BSDE part of the scheme. Note that there is no contradiction with what is predicted by Theorem 4.3. Indeed, it is expected that for very large values of N the asymptotic convergence will begin to take place.[15]

For bigger values of α (between 40 and 60), we can finally observe the transition to the asymptotic regime happening in our window of N's.

Finally, for larger values of α (60 and above), we mark on Figure 2(b) only the finite values of $e(N)$ [defined in (7.1)]. This shows in a rather clear fashion that if we do not truncate strongly enough (for a given value of N) the scheme "blows up" (the code produces *NaN* values). One also observes that the bigger the multiplying factor α the smaller the time-step must be in order to make sure that $e(N)$ decays appropriately (converges). This depicts very well the scenario described in our counter-example. We believe that the high convergence rates appearing in Table 2 when α is big is due to the smoothness of the driver f we chose for Example 2 (similar to Example 1) and its damping effect on the dynamics of the scheme. We leave an in-depth analysis of this fact for future research.

[15]In order to significantly increase N, we would also need to increase M to levels that are beyond our computational capabilities.

APPENDIX

A.1. Motivating example. Before we state the main result, we recall a result on the behavior of Gaussian random variables [which we do not prove, but the reader is invited to try, in any case see Lemma 4.1 in Hutzenthaler, Jentzen and Kloeden (2011)]. The notation and probability spaces we work with in this Appendix are as stated in Section 2.

LEMMA A.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $Z : \Omega \to \mathbb{R}$ be an $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable mapping with standard normal distribution. Then for any $x \in [0, \infty)$ it holds that*

$$\mathbb{P}\big[|Z| \geq x\big] \geq \tfrac{1}{4} x e^{-x^2}.$$

The statement of Lemma 1.1 follows from the next lemma.

LEMMA A.2. *Let π^N denote the uniform grid of the time interval $[0, 1]$ with $N+1$ points and step size $h := 1/N$, where $N \in \mathbb{N}$. Define the driver $f(y) := -y^3$ and the terminal condition $\xi \in L^p(\mathcal{F}_1)$ for any $p \geq 2$. Let (Y, Z) be the unique solution to (1.3). Denote by $\{Y_i^{(N)}\}_{i \in \{0,\dots,N\}}$ the Euler approximation of $(Y_t)_{t \in [0,1]}$ defined via (1.4) over the grid π^N.*

Assume that N is fixed and that ξ verifies $|\xi| \geq 2\sqrt{N}$ $\mathbb{P}$-a.s. then:

(i) *For any $i \in \{0, \dots, N\}$ it holds that $|Y_i| \geq 2^{2^{N-i}}\sqrt{N}$.*

Assume now that N is an even *number (hence $t = 1/2$ is common to all grids π^N) and denote by $Y_{1/2}^{(N)}$ the approximation at the time point $t = 1/2$ (corresponding to $i = N/2$). Define ξ as $\xi := W_{1/2} \in L^p(\mathcal{F}_1) \setminus L^\infty(\mathcal{F}_1)$ for any $p \geq 1$.*

(ii) *For any $i \in \{\frac{N}{2}, \dots, N\}$, on the set $\{\omega : \xi(\omega) \geq 2\sqrt{N}\}$ it holds that $|Y_i(\omega)| \geq 2^{2^{N-i}}\sqrt{N}$.*

(iii) *Moreover, $\lim_{N\to\infty} \mathbb{E}[|Y_{1/2}^{(N)}|] = +\infty$.*

PROOF. For the given f and ξ, the results from Section 2 in Pardoux (1999) combined with the a priori estimates stated in our Section 2 ensure the existence and uniqueness of a solution $(Y, Z) \in \mathcal{S}^p \times \mathcal{H}^p$ to BSDE (1.3) for any $p \geq 2$. We now fix N and drop the superscript (N) from $Y^{(N)}$.

Proof of Part (i). Without loss of generality, assume that $\xi = Y_N \geq 2\sqrt{N}$. Then

$$Y_{N-1} = \mathbb{E}_{N-1}\big[Y_N - Y_N^3 h\big] = \mathbb{E}_{N-1}\big[Y_N\big(1 - Y_N^2 h\big)\big].$$

Observe that $Y_N^2 \geq 2N$ which implies $(1 - Y_N^2 h) \leq (1 - 2^2) < 0$. Hence (since $Y_N > 0$),

$$Y_{N-1} = \mathbb{E}_i\big[Y_N\big(1 - Y_N^2 h\big)\big] \leq -2\sqrt{N}(2^2 - 1) \leq -2^2\sqrt{N}.$$

Next (since $Y_{N-1} < 0$) $Y_{N-1}^2 \geq 2^4 N$ which implies $1 - Y_{N-1}^2 h \leq (1 - 2^4) < 0$. Hence,

$$Y_{N-2} = \mathbb{E}_i\big[Y_{N-1}\big(1 - Y_{N-1}^2 h\big)\big] = \mathbb{E}_i\big[(-Y_{N-1})\big(Y_{N-1}^2 h - 1\big)\big]$$
$$\geq 2^2\sqrt{N}\big(2^4 - 1\big) \geq 2^{2^2}\sqrt{N}.$$

Proceeding by induction, we can show that

$$|Y_i| \geq 2^{2^{N-i}}\sqrt{N}.$$

Indeed, assume $|Y_{i+1}| \geq 2^{2^{N-i-1}}\sqrt{N}$ (in the light of above calculations; the negative case is analogous), then

$$Y_i = \mathbb{E}_i\big[Y_N\big(1 - Y_N^2 h\big)\big] \leq 2^{2^{N-i-1}}\sqrt{N}\big(\big(2^{2^{N-i-1}}\big)^2 - 1\big) \leq 2^{2^{N-i}}\sqrt{N}$$

and statement (i) is proved.

Before proving (ii) and (iii), we remark that no conditional expectation needs to be computed for the scheme (1.4) for $i \in \{N/2, \dots, N\}$ because $\xi = W_{1/2}$ is $\mathcal{F}_t$-adapted for any $t \in [1/2, 1]$. The scheme's approximations up to $Y_{1/2}^{(N)}$ can be written as

$$Y_N^{(N)} = W_{1/2}, \qquad Y_{N-1}^{(N)} = \psi(W_{1/2}),$$
$$Y_{N-2}^{(N)} = \psi\big(\psi(W_{1/2})\big), \qquad \dots, \qquad Y_{N/2}^{(N)} = \psi^{\circ(N/2)}(W_{1/2}),$$

where $\psi(x) := x - hx^3$ and $\psi^{\circ(n)}$ denotes the composition of ψ with itself n-times ($n \in \mathbb{N}$).

Proof of Part (ii). We work on the event that $\xi = Y_N \geq 2\sqrt{N}$. We have first

$$Y_{N-1} = \mathbb{E}_{N-1}\big[Y_N - Y_N^3 h\big] = Y_N\big(1 - Y_N^2 h\big).$$

Observe that $Y_N^2 \geq 2^2 N$ which implies $(1 - Y_N^2 h) \leq (1 - 2^2) < 0$. Hence (since $Y_N > 0$),

$$Y_{N-1} = Y_N\big(1 - Y_N^2 h\big) \leq -2\sqrt{N}\big(2^2 - 1\big) \leq -2^2\sqrt{N} < 0.$$

Next, since $Y_{N-1} < 0$, $Y_{N-1}^2 \geq 2^4 N$ which implies $1 - Y_{N-1}^2 h \leq (1 - 2^4) < 0$. Hence,

$$Y_{N-2} = Y_{N-1}\big(1 - Y_{N-1}^2 h\big) = -Y_{N-1}\big(Y_{N-1}^2 h - 1\big) \geq 2^2\sqrt{N}\big(2^4 - 1\big) \geq 2^{2^2}\sqrt{N}.$$

Proceeding by induction we can easily show that

$$|Y_i| \geq 2^{2^{N-i}}\sqrt{N}, \qquad i = \frac{N}{2}, \dots, N.$$

Indeed, assume $Y_{i+1} \geq 2^{2^{N-i-1}}\sqrt{N}$ (note that in the light of the above calculations the negative case is analogous). Then

$$Y_i = Y_{i+1}\big(1 - Y_{i+1}^2 h\big) \leq 2^{2^{N-i-1}}\sqrt{N}\big(1 - \big(2^{2^{N-i-1}}\big)^2\big) \leq -2^{2^{N-i}}\sqrt{N}.$$

Proof of Part (iii). It follows easily from Lemma A.1 that

$$\mathbb{P}[|W_{1/2}| \geq 2\sqrt{N}] \geq \frac{\sqrt{2}}{2}\sqrt{N}e^{-8N}.$$

Then, using part (i) (to go from the first to the second line) and the above remark (on the third line), we have

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{E}[|Y_{1/2}^{(N)}|] \\
&\quad= \lim_{N\to\infty} \mathbb{E}[\mathbb{1}_{\{\xi \geq 2\sqrt{N}\}}|Y_{1/2}^{(N)}| + \mathbb{1}_{\{\xi < 2\sqrt{N}\}}|Y_{1/2}^{(N)}|] \\
&\quad\geq \lim_{N\to\infty} \mathbb{E}[\mathbb{1}_{\{\xi \geq 2\sqrt{N}\}}|Y_{1/2}^{(N)}|] \\
&\quad\geq \lim_{N\to\infty} \mathbb{E}[\mathbb{1}_{\{\xi \geq 2\sqrt{N}\}} 2^{2^{N-N/2}}\sqrt{N}] \\
&\quad= \lim_{N\to\infty} 2^{2^{N/2}}\sqrt{N}\mathbb{P}[|W_{1/2}| \geq 2\sqrt{N}] \\
&\quad\geq \lim_{N\to\infty} 2^{(2^{N/2})}\frac{\sqrt{2}}{2}Ne^{-8N} = +\infty.
\end{aligned}$$

□

A.2. Basics of Malliavin's calculus. We briefly introduce the main notation of the stochastic calculus of variations also known as Malliavin's calculus. For more details, we refer the reader to Nualart (2006), for its application to BSDEs we refer to Imkeller (2008). Let $\boldsymbol{\mathcal{S}}$ be the space of random variables of the form

$$\xi = F\left(\left(\int_0^T h_s^{1,i}\,\mathrm{d}W_s^1\right)_{1\leq i\leq n}, \dots, \left(\int_0^T h_s^{d,i}\,\mathrm{d}W_s^d\right)_{1\leq i\leq n}\right),$$

where $F \in C_b^\infty(\mathbb{R}^{n\times d})$, $h^1, \dots, h^n \in L^2([0,T];\mathbb{R}^d)$, $n \in \mathbb{N}$. To simplify notation, assume that all h^j are written as row vectors. For $\xi \in \boldsymbol{\mathcal{S}}$, we define $D = (D^1, \dots, D^d) : \boldsymbol{\mathcal{S}} \to L^2(\Omega \times [0,T])^d$ by

$$D_\theta^i \xi = \sum_{j=1}^n \frac{\partial F}{\partial x_{i,j}}\left(\int_0^T h_t^1\,\mathrm{d}W_t, \dots, \int_0^T h_t^n\,\mathrm{d}W_t\right)h_\theta^{i,j}, \qquad 0 \leq \theta \leq T, 1 \leq i \leq d,$$

and for $k \in \mathbb{N}$ its k-fold iteration by $D^{(k)} = (D^{i_1} \cdots D^{i_k})_{1\leq i_1,\dots,i_k\leq d}$. For $k \in \mathbb{N}$, $p \geq 1$ let $\mathbb{D}^{k,p}$ be the closure of $\mathcal{S}$ with respect to the norm

$$\|\xi\|_{k,p}^p = \mathbb{E}\left[\|\xi\|_{L^p}^p + \sum_{i=1}^k \||D^{(k)}\xi|\|_{(\mathcal{H}^p)^i}^p\right].$$

$D^{(k)}$ is a closed linear operator on the space $\mathbb{D}^{k,p}$. Observe that if $\xi \in \mathbb{D}^{1,2}$ is $\mathcal{F}_t$-measurable then $D_\theta\xi = 0$ for $\theta \in (t,T]$. Further denote $\mathbb{D}^{k,\infty} = \bigcap_{p>1}\mathbb{D}^{k,p}$.

We also need Malliavin's calculus for $\mathbb{R}^m$ valued smooth stochastic processes. For $k \in \mathbb{N}$, $p \geq 1$, denote by $\mathbb{L}^{k,p}(\mathbb{R}^m)$ the set of $\mathbb{R}^m$-valued progressively measurable processes $u = (u^1, \dots, u^m)$ on $[0,T] \times \Omega$ such that:

(i) For Lebesgue-a.a. $t \in [0,T]$, $u(t,\cdot) \in (\mathbb{D}^{k,p})^m$;

(ii) $[0,T] \times \Omega \ni (t,\omega) \mapsto D^{(k)}u(t,\omega) \in (L^2([0,T]^{1+k}))^{d\times n}$ admits a progressively measurable version;

(iii) $\|u\|_{k,p}^p = \|u\|_{\mathcal{H}^p}^p + \sum_{i=1}^k \|D^i u\|_{(\mathcal{H}^p)^{1+i}}^p < \infty$.

Note that Jensen's inequality gives[16] for all $p \geq 2$

$$\mathbb{E}\bigg[\bigg(\int_0^T \int_0^T |D_u X_t|^2 \,\mathrm{d}u \,\mathrm{d}t\bigg)^{p/2}\bigg] \leq T^{p/2-1} \int_0^T \|D_u X\|_{\mathcal{H}^p}^p \,\mathrm{d}u. \tag{A.1}$$

We recall a result from Imkeller (2008) concerning the rule for the Malliavin differentiation of Itô integrals which is of use in applications of Malliavin's calculus to stochastic analysis.

THEOREM A.3 [Theorem 2.3.4 in Imkeller (2008)]. *Let $(X_t)_{t\in[0,T]} \in \mathcal{H}^2$ be an adapted process and define $M_t := \int_0^t X_r \,\mathrm{d}W_r$ for $t \in [0,T]$. Then $X \in \mathbb{L}^{1,2}$ if and only if $M_t \in \mathbb{D}^{1,2}$ for any $t \in [0,T]$.*

Moreover, for any $0 \leq s,t \leq T$ we have

$$D_s M_t = X_s \mathbb{1}_{\{s\leq t\}}(s) + \mathbb{1}_{\{s\leq t\}}(s) \int_s^t D_s X_r \,\mathrm{d}W_r. \tag{A.2}$$

A.3. A particular Gronwall lemma. We state here a "discrete Gronwall lemma" of some kind, particularly useful for the numerical analysis of BSDEs, and which we use extensively in this work.

LEMMA A.4. *Let a_i, b_i, c_i, be such that $a_i, b_i \geq 0$, $c_i \in \mathbb{R}$ for $i = 0,1,\ldots,N$. Assume that, for some constant $c > 0$ and $h > 0$, we have*

$$a_i + b_i \leq (1+ch)a_{i+1} + c_i \qquad \text{for } i = 0,1,\ldots,N-1. \tag{A.3}$$

Then the following inequality holds for every i:

$$a_i + \sum_{j=i}^{N-1} b_j \leq e^{c(N-i)h} a_N + \sum_{j=i}^{N-1} e^{c(j-i)h} c_j.$$

PROOF. The estimate is clearly true for $i = N-1$ (even for $i = N$ in fact). Then, for any $i \leq N-2$, if it is true for $i+1$, by multiplying both sides by e^{ch} we find that

$$e^{ch} a_{i+1} + e^{ch} \sum_{j=i+1}^{N-1} b_j \leq e^{c(N-i)h} a_N + \sum_{j=i+1}^{N-1} e^{c(j-i)h} c_j.$$

Summing this inequality with (A.3) and noting that $\sum_{j=i+1}^{N-1} b_j \leq e^{ch} \sum_{j=i+1}^{N-1} b_j$ due to the positivity of the b_j terms gives the sought estimate for any i. □

[16]The reason behind this last inequality is that within the BSDE framework the usual tools to obtain a priori estimates yield with much difficulty the LHS while with relative ease the RHS.

Acknowledgments. We would like to thank Samuel Cohen (University of Oxford), Gechun Liang (King's college) and Joscha Diehl (TU-Berlin) for helpful discussions.

REFERENCES

ALANKO, S. and AVELLANEDA, M. (2013). Reducing variance in the numerical solution of BSDEs. *C. R. Math. Acad. Sci. Paris* **351** 135–138. MR3038003

BOUCHARD, B. and TOUZI, N. (2004). Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations. *Stochastic Process. Appl.* **111** 175–206. MR2056536

BRIAND, P. and CARMONA, R. (2000). BSDEs with polynomial growth generators. *J. Appl. Math. Stoch. Anal.* **13** 207–238. MR1782682

BRIAND, P. and CONFORTOLA, F. (2008). Differentiability of backward stochastic differential equations in Hilbert spaces with monotone generators. *Appl. Math. Optim.* **57** 149–176. MR2386102

BRIAND, PH., DELYON, B., HU, Y., PARDOUX, E. and STOICA, L. (2003). L^p solutions of backward stochastic differential equations. *Stochastic Process. Appl.* **108** 109–129. MR2008603

CHASSAGNEUX, J. F. (2012). An introduction to the numerical approximation of BSDEs. Lecture notes, 2nd Summer School of the Euro-Mediterranean Research Center for Mathematics and its Applications (EMRCMA). Available at www.imperial.ac.uk/~jchassag/.

CHASSAGNEUX, J. F. (2013). Linear multi-step schemes for BSDEs. Preprint. Available at arXiv:1306.5548v1.

CHASSAGNEUX, J. F. and CRISAN, D. (2014). Runge–Kutta schemes for backward stochastic differential equations. *Ann. Appl. Probab.* **24** 679–720. MR3178495

CHASSAGNEUX, J.-F. and RICHOU, A. (2013). Numerical simulation of quadratic BSDEs. Preprint. Available at arXiv:1307.5741.

CRISAN, D. and MANOLARAKIS, K. (2010). Second order discretization of backward SDEs and simulation with the cubature method. *Ann. Appl. Probab.* **24** 652–678. MR3178494

CRISAN, D. and MANOLARAKIS, K. (2012). Solving backward stochastic differential equations using the cubature method: Application to nonlinear pricing. *SIAM J. Financial Math.* **3** 534–571. MR2968045

DOS REIS, G., RÉVEILLAC, A. and ZHANG, J. (2011). FBSDEs with time delayed generators: L^p-solutions, differentiability, representation formulas and path regularity. *Stochastic Process. Appl.* **121** 2114–2150. MR2819244

EL KAROUI, N., PENG, S. and QUENEZ, M. C. (1997). Backward stochastic differential equations in finance. *Math. Finance* **7** 1–71. MR1434407

ESTEP, D. J., LARSON, M. G. and WILLIAMS, R. D. (2000). Estimating the error of numerical solutions of systems of reaction–diffusion equations. *Mem. Amer. Math. Soc.* **146** viii+109. MR1692630

GOBET, E., LEMOR, J.-P. and WARIN, X. (2005). A regression-based Monte Carlo method to solve backward stochastic differential equations. *Ann. Appl. Probab.* **15** 2172–2202. MR2152657

GOBET, E. and TURKEDJIEV, P. (2011). Approximation of discrete BSDE using least-squares regression. Technical Report hal-00642685.

HENRY, D. (1981). *Geometric Theory of Semilinear Parabolic Equations*. *Lecture Notes in Math.* **840**. Springer, Berlin. MR0610244

HUTZENTHALER, M., JENTZEN, A. and KLOEDEN, P. E. (2011). Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **467** 1563–1576. MR2795791

HUTZENTHALER, M., JENTZEN, A. and KLOEDEN, P. E. (2012). Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *Ann. Appl. Probab.* **22** 1611–1641. MR2985171

IMKELLER, P. (2008). *Malliavin's Calculus and Applications in Stochastic Control and Finance*. *IMPAN Lecture Notes* **1**. Polish Academy of Sciences, Institute of Mathematics, Warsaw. MR2494786

IMKELLER, P. and DOS REIS, G. (2010a). Path regularity and explicit convergence rate for BSDE with truncated quadratic growth. *Stochastic Process*. *Appl*. **120** 348–379. MR2584898

IMKELLER, P. and DOS REIS, G. (2010b). Corrigendum to "Path regularity and explicit convergence rate for BSDE with truncated quadratic growth" [Stochastic Process. Appl. **120** (2010) 348–379] [MR2584898]. *Stochastic Process*. *Appl*. **120** 2286–2288.

KLOEDEN, P. E. and PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations*. *Applications of Mathematics* (*New York*) **23**. Springer, Berlin. MR1214374

KOVÁCS, B. (2011). Semilinear parabolic problems. Master's thesis, Eötvös Loránd Univ., Budapest.

LIONNET, A. (2014). Topics on backward stochastic differential equations. Theoretical and practical aspects. Ph.D. thesis, Oxford Univ.

MA, J. and ZHANG, J. (2002). Path regularity for solutions of backward stochastic differential equations. *Probab*. *Theory Related Fields* **122** 163–190. MR1894066

MAO, X. and SZPRUCH, L. (2013). Strong convergence rates for backward Euler–Maruyama method for non-linear dissipative-type stochastic differential equations with super-linear diffusion coefficients. *Stochastics* **85** 144–171. MR3011916

MATOUSSI, A. and XU, M. (2008). Sobolev solution for semilinear PDE with obstacle under monotonicity condition. *Electron*. *J*. *Probab*. **13** 1035–1067. MR2424986

MILSTEIN, G. N. and TRETYAKOV, M. V. (2004). *Stochastic Numerics for Mathematical Physics*. *Scientific Computation*. Springer, Berlin. MR2069903

NUALART, D. (2006). *The Malliavin Calculus and Related Topics*, 2nd ed. Springer, Berlin. MR2200233

PARDOUX, É. (1999). BSDEs, weak convergence and homogenization of semilinear PDEs. In *Nonlinear Analysis*, *Differential Equations and Control* (*Montreal*, *QC*, 1998). *NATO Sci*. *Ser. C Math*. *Phys*. *Sci*. **528** 503–549. Kluwer Academic, Dordrecht. MR1695013

ROTHE, F. (1984). *Global Solutions of Reaction-Diffusion Systems*. *Lecture Notes in Math*. **1072**. Springer, Berlin. MR0755878

SÜLI, E. and MAYERS, D. F. (2003). *An Introduction to Numerical Analysis*. Cambridge Univ. Press, Cambridge. MR2006500

TOUZI, N. (2013). *Optimal Stochastic Control*, *Stochastic Target Problems*, *and Backward SDE*. *Fields Institute Monographs* **29**. Springer, New York. MR2976505

ZEIDLER, E. (1990). *Nonlinear Functional Analysis and Its Applications*. *II/B*. *Nonlinear Monotone Operators*. Springer, New York. Translated from the German by the author and Leo F. Boron. MR1033498

ZHANG, G., GUNZBURGER, M. and ZHAO, W. (2013). A sparse-grid method for multi-dimensional backward stochastic differential equations. *J*. *Comput*. *Math*. **31** 221–248. MR3063734

ZHANG, Q. and ZHAO, H. (2012). Probabilistic representation of weak solutions of partial differential equations with polynomial growth coefficients. *J*. *Theoret*. *Probab*. **25** 396–423. MR2914434

A. LIONNET
OXFORD-MAN INSTITUTE
UNIVERSITY OF OXFORD
EAGLE HAUSE, WALTON WELL ROAD
OXFORD OX2 6ED
UNITED KINGDOM
E-MAIL: arnaud.lionnet@maths.ox.ac.uk

G. DOS REIS
L. SZPRUCH
SCHOOL OF MATHEMATICS
UNIVERSITY OF EDINBURGH
EDINBURGH, EH9 3JZ
UNITED KINGDOM
E-MAIL: G.dosReis@ed.ac.uk
l.szpruch@ed.ac.uk